Reservoir SMILES: Towards SensoriMotor Interaction of Language and Embodiment of Symbols with Reservoir Architectures

Xavier Hinaut

UNIVERSITY OF BORDEAUX

DOCTORAL SCHOOL EDMI

H A B I L I T A T I O N

of the University of Bordeaux

Specialty: Computer Science

Defended by

Xavier HINAUT

Reservoir SMILES

Towards SensoriMotor Interaction of Language and Embodiment of Symbols with Reservoir Architectures

Prepared at Inria Bordeaux / LaBRI / IMN, Mnemosyne Team

defended on November 16th, 2022

at New-Aquitaine **Broca** Center

Jury:

Reviewers:		
Dr. Julien Diard	-	CNRS, Grenoble, FR
Prof. Joanna Rączaszek-Leonardi	-	University of Warsaw, PL
Prof. Tadahiro Taniguchi	-	Ritsumeikan University, JP
Examinators:		
Prof. Morten Christiansen	-	Cornell University, USA
Prof. Abigail Morrison	-	Jülich Supercomputing Centre, DE
Dr. Alexandre Pitti	-	CY Cergy Paris Université, FR
President:		
Dr. Christophe Pallier	-	INSERM-CEA, Neurospin, FR

A toutes celles et ceux qui me soutiennent de façon visible ou invisible.[1]

Ce qui nous est invisible met plus de temps à être étudié.
Mais en parler le fait exister.
On en vient à conceptualiser l'invisible.[2] [3]

[1] To all those who support me in a visible or invisible way.

[2] What is invisible to us takes longer to be studied. But talking about it makes it exist. We come to conceptualize the invisible.

[3] I wrote these sentences before the Nobel Prize of Physics was revealed on October 4th, but of course quantum entanglement is a good example of it.

Acknowledgements

I first want to thank the jury members, Morten Christiansen, Julien Diard, Abigail Morrison, Christophe Pallier, Alexandre Pitti, Joanna Rączaszek-Leonardi, Tadahiro Taniguchi, to have accepted to take part in my Habilitation defense, and especially the official reviewers who have the difficult task to summarize what I have already tried to summarize myself.

I warmly thank all the actual and past members of the Mnemosyne team – along with IMN and Inria members – for the great work atmosphere and vibrant discussions in or outside the lab! I would like to thank them individually, but I am too afraid of forgetting someone and/or all the reasons I want to say thank you. I am grateful to Frédéric Alexandre for making the existence of this great team possible and having enabled me to integrate it since more than six years ago.

Of course I also equivalently thank my previous colleagues from Lyon (Peter Dominey, Pierre Enel, Anne-Lise Jouen, Stéphane Lallée, ...), Orsay/Saclay (Catherine Del Negro, Nicolas Giret, Aurore Cazala, Christophe Pallier, ...), and Hamburg (Stefan Wermter, Johannes Twiefel, Nicolas Navarro-Guerrero, Jorge Davillas-Chacon, Carina Gaber, Doreen Jirak, ...) because many discussions influence my research today.

I have to make a special thanks to Peter Dominey and Catherine Del Negro who "branched" my research on language and songbirds respectively, without whom this may not have happened. Particular thanks to Peter who unintentionally inspired my later collaboration with Stefan Franck (who I also thank) by producing code-switching sentences in "Frenglish".

A big thanks to Anthony and Silvia, my previous PhD students, and Subba, my current PhD student, with whom I had very rich and passionate discussions. I thank all the interns who worked with me – directly or indirectly –, for their work, our discussions, ideas, ... and without whom I could not have done all theses studies. Of course, I am also grateful to the co-supervisors of these PhD students for their trust and patience and enabling me to have had my first experiences in managing these great scientific minds: Nicolas Rougier, Arthur Leblois and Frédéric Alexandre. I am obliged to Nathan Trouvain for his nice work on ReservoirPy, his help on several projects, and his motivation to start his PhD on this ambitious research program.

I warmly thank Gaël Jobard for all the interesting and very inspiring exchanges we had on language, some of which helped me define my research program, for the proofreading of earlier versions of this research program, and for the collaboration we started. I say thank you to Catherine Del Negro and Aurore Cazala for their nice Marseille "southness" and all the time we spent to "mine" for chunks in the canary song wild.

I have to thank the authors of some very inspiring books that still trigger me towards far-fetch scientific goals, in chronological order: *Why red does not sound like a bell?* from Kevin O'Regan [O'Regan 2011], *La Révolution de l'Intelligence du Corps* from Alex Pitti and Rolf Pfeiffer [Pfeifer & Pitti 2012] and *Creating language: Integrating evolution, acquisition, and processing* from Morten Christiansen

and Nick Chater [Christiansen *et al.* 2016].

I would like to show my gratitude to several people with whom I had interesting exchanges in Bordeaux or elsewhere; here is an incomplete list: Mehdi Khamassi for scientific and non scientific exchanges, and for his advice over the years (already before the starting of my PhD); Kevin O'Regan[4] and Lisa Jacquet[5] for the very nice and refreshing discussions and karaoke at ICDL conferences; Justine Mertz for sharing her inspiring knowledge on sign languages; and many others ...

I express my gratitude to Clément Moulin-Frier for the mutual support and his advice during the writing of this Habitation Thesis, for co-organising this great SMILES workshop with me for almost 3 years now, for his work inspiring my research program and of course for the long scientific and non scientific exchanges. Moreover, I am obliged to Nicolas Rougier for offering his help and beautiful template for my list of publications. Finally, I am very grateful to Chloé, Clara, Clément, Hippo, Hugo and Nathan for proofreading this manuscript and for their helpful comments!

I thank all the people that helped me organise the *Hack'1 Cerveau* Hackathon[6] in December 2017 at Cap Sciences museum, along with the participants, the members of the jury and the partners that made it possible: after participating in several hackathons, I was eager to find the opportunity to organise myself one.

I'd like to thank teachers of the University of Technology of Compiègne and members of the Akureyri seminar in Iceland, which initiated early on (2005-2006) my brain to fundamental cognitive science and philosophical thinking such as child development, enaction or autopoiesis. I thank Jean-Philippe Rennard for his popularization books and website on Artificial Neural Networks and Artificial Life, who helped me assuage my desire for biological modelling when Youtube was not full of youtubers. Although it's not categorized as science, I should thank Bernard Werber for his novel *L'ultime secret*[7], which is probably one of the strong "kernels" that kept pushing my interests towards brain understanding until discovering the reservoir "trick". I thank Theo Jansen, with whom I had the chance to discuss with recently, for his incredible creatures that really *feel like* animals walking when you watch them.

I give thanks to my musical colleagues with whom I had the chance to improvise with my saxophone. In chronological order: Anthony, Basile, Aurélien, Hugo (and Chloé who helped catching melody pitches). It has been better than team building. Improvising enables to let your artistic creativity free, which in turn probably enriches your scientific one. I am obliged to my teacher Julien for pushing me enhancing my saxophone skills and Arthur Leblois for lending me his saxophone at the beginning. That's why I now want to show my gratitude to people with whom I have explored another form of creativity: Pascal Etcheto from the University of

[4] And thanks for his declamation of the Jabberwocky from Lewis Carroll.

[5] And thanks for diving me in her creative projects.

[6] https://pixees.fr/hack1-cerveau/

[7] Although I am fascinated by several Isaac Asimov' books, in particular for his scientific and historical way of analysing things and for being able to lately anchor chaos theory to his odyssey, I was already working on AI when I did read them.

Bordeaux, the groups of the contemporary dance classes, teachers and members of Bordeaux Swing and Dance in West, and of course all previous teachers and dancers in Lyon, Paris and Orsay. Besides, I also want to thank my surf teachers (which I started more recently) for their patience and advice to learn how to be *synchronized* or *coupled* with the waves. Surf classes, alike contemporary dance, both sometimes helped me to disconnect from everyday life by connecting to the wild, and sometimes helped me to be very focused on problems and find new ideas. For similar reasons, I should thank again Catherine for offering me this postdoc position on songbirds which triggered a strange habit: a month after I started I could no longer ignore birdsongs why walking in the street. Now that I am able to recognize more than a dozen, it is easier to disconnect anywhere where there are birds.

I am thankful to all the musical artists that made songs that motivated me and kept me focusing during the writing of this manuscript, in particular Led Zeppelin who helped again like during my PhD manuscript writing, as well as other old rock bands: UFO[8], Cream[9], Lynyrd Skynyrd[10], Mountain, Jimmy Hendrix[11], Deep Purple[12], Scorpions, Pink Floyd [13], and many others[14]; Of course, I am also grateful to electronic musicians which also helped me writing: Fatboy Slim[15], The Chemical Brothers[16], Jean du voyage, Daft Punk, Wax Tailor[17], Chinese Man, Amon Tobin, The Prodigy, Soulwax, Gabriel Rios[18], ...; and also funk music like Chaka Khan[19], and other artists I can't classify easily like Chapelier Fou, Grand National [20], Jain ... And last but not least, *Metronomy* for their song that helps dive into reservoirs.

I thank Clara and Hippo for the mix of scientific discussions, advice and crazy trips we had in the last years, which helped me both scientifically and personally before, during and after these strange COVID times. I grasp the opportunity to thank the past members and friends of the French young cognitive science community (Fresco et al.) for all these great moments and endless chats on original topics.

I thank my family and close friends for they lifelong support and because they are the early and important influencers of my life – which include my scientific ones – especially by making me discover early on artistic and scientific material.

Finally I thank all the "invisible" people I can't remember at the precise moment of writing these lines, and people that helped me change my brain dynamics with

[8] *Rock bottom.*

[9] *Sunshine of your love* & *White room* that are especially motivating.

[10] *Free Bird.*

[11] *Purple Haze.*

[12] *Smoke on the water.*

[13] *Shine on you crazy diamond (Live).*

[14] Maybe as some point I should study why old rock music helps my brain focus while writing.

[15] *Ya Mamma*, *Right here, right now.*

[16] *Galvanize*, *Hey Boy Hey Girl.*

[17] It is easy to follow all the tracks from an album in a row as it tells a story. Special thank for his *Que sera* song which followed me since my first cognitive science courses.

[18] *Broad Daylight.*

[19] *Like Sugar.*

[20] *Peanut Dreams.*

non-scientific exchanges which had an impact on my scientific thoughts.

Because we are not – and *never* will be – "brains in vats", but body-brains interacting with the world, my current research would not have been the same without all these people.

Contents

Reservoir SMILES: Towards SensoriMotor Interaction of Language and Embodiment of Symbols with Reservoir Architectures

Abstract: Language involves several hierarchical levels of abstraction. Most models focus on a particular level of abstraction making them unable to model bottom-up and top-down processes. Moreover, we do not know how the brain grounds symbols to perceptions and how these symbols emerge throughout development. Experimental evidence suggests that perception and action shape one-another (e.g. motor areas activated during speech perception) but the precise mechanisms involved in this action-perception shaping at various levels of abstraction are still largely unknown.

My previous and current work include the modelling of language comprehension, language acquisition with a robotic perspective, sensorimotor models and extended models of Reservoir Computing to model working memory and hierarchical processing. I propose to create a new generation of neural-based computational models of language processing and production; to use biologically plausible learning mechanisms relying on recurrent neural networks; create novel sensorimotor mechanisms to account for action-perception shaping; build hierarchical models from sensorimotor to sentence level; embody such models in robots.

Keywords: Reservoir Computing, Echo State Networks, Language Processing, Language Acquisition, Songbird, Sound Classification, Sound Generation, SensoriMotor, Action-Perception, Model, Robot, Sequences, Chunking, Symbol Emergence, Symbol Grounding Problem, Computational Neuroscience

List of Acronyms

Chapter 1

Introduction

It is remarkable how from earliest forms of life some species evolved to produce complex sequences of symbols using air or water vibrations. Concerning humans, how languages started and evolved is an intriguing question, since it is difficult to collect evidence before the invention of writing. This evolutionary perspective is important to keep in mind when studying language, even if we tend to forget it when we put language and its substrates (humans, books, ...) under scientific controlled conditions. Studying language at the individual level, in relation with what happens in the brain of this individual, is already a complex enough task. My works focused at this level so far.

Syntax, one of the most abstract parts of language, is often analysed in isolation of individuals, their brains, their bodies and their environment. But could we really ignore the influence of all these things? Let's consider how sign languages deal with anaphoric references. An anaphoric reference is the use of a term that is referring back to something said previously. For instance, let's consider the sentence: "The *cat* sat on the mat. *She* ate all the seeds while the birds where delighted." *She* is an anaphoric reference to the *cat*. In sign languages, "cat" sign would be associated to a particular position in space, and the anaphoric reference would be obtained by pointing towards that position in space [Schlenker 2017]. Thus in sign languages, sentences are not only a linear sequence of symbols[1]. They use physical space to anchor some particular elements in a stream of symbols. Whereas spoken languages can not use space to "extract" some elements from the linear stream of an utterance. Besides, it could be considered that syntax is not only present in languages, but also in actions. Pulvermüller talks about *syntax of actions* [Pulvermüller 2014], although it is a matter of debate [Moro 2014]. [Pastra & Aloimonos 2012] proposed a *minimalist grammar of actions* inspired from linguistics analysis tools.

More generally, cortical areas often associated with syntax processing are not specific to language processing[2]. For instance, Broca area – more specifically the Left Inferior Frontal Gyrus (LIFG) – is not only involved with syntax processing, but has been found to be related to action recognition and movement preparation [Thoenissen *et al.* 2002, Hamzei *et al.* 2003, Hagoort 2005]. This link between language and actions in Broca area, together with its involvement in the primate "Mirror Neuron System (MNS)", suggests how primate brains may

[1] Of course, in spoken languages some of these symbols could be stressed by prosody.

[2] As we will see in Chapter 3 (Subsection 3.2) reservoirs are interesting in this respect because they are not specialized to a specific function: a single reservoir can be used for several independent tasks.

have evolved from action to language representations [Rizzolatti & Arbib 1998, Fadiga *et al.* 2009]. Similar kind of neurons have been found in a sensorimotor area of songbirds[3] [Prather *et al.* 2008]. Broca area has also been shown to be involved in music execution and listening [Zatorre *et al.* 2007] suggesting that it is more generally involved in representing hierarchical-like structures [Koechlin & Jubault 2006, Fadiga *et al.* 2009].

The brain is hierarchically organized from more perceptual to more integrative areas [Felleman & Van Essen 1991, Markov *et al.* 2013]. Understanding how hierarchical processes are organised [Koechlin & Jubault 2006] and modelling such processes in language and other modalities are part of the long-term goals of my work. Since my PhD[4], I am interested in these questions of hierarchical organisation and hierarchical processes [Markov & Kennedy 2013]. Under the supervision of Peter Dominey, I started my PhD by making a three-layered model of primate cortex encoding categories of sequences [Hinaut & Dominey 2011]. Then, we used more general reservoirs to model Broca area with respect to grammatical constructions[5] processing. The model was able to generalize to unseen constructions, learn anaphoric references and make role predictions of upcoming words during sentence parsing [Hinaut & Dominey 2012, Hinaut & Dominey 2013]. This model was embedded into the humanoid iCub robot: we could teach it to link sequences of actions to grammatical constructions[6], and vice versa (i.e. the robot produced sentences)[7] [Hinaut *et al.* 2014]. Shortly after my PhD, we managed to have the model to learn incrementally with an online learning rule [Hinaut & Wermter 2014]. We also studied the generalization capabilities of the model for production of sentences [Hinaut *et al.* 2015a]. Afterwards, we made the sentence processing more robust: it was able to process Out-of-Vocabulary (OOV) words [Hinaut *et al.* 2015b], work with different sentence levels of abstraction (sequences of phonemes, of words or grammatical constructions) [Hinaut 2018], and use various kinds of meaning representations [Hinaut *et al.* 2016]. With collaborators at Hamburg University, we embedded the model into the humanoid Nao robot to learn new objects with depth camera and robust speech recognition[8] [Hinaut *et al.* 2015c]: it was robust enough to work in a noisy environment such as the Science Night popularization event. Moreover, we managed to train a single reservoir to learn two languages [Hinaut *et al.* 2015b] and to understand "code-switched" bilingual sentences[9] [Detraz & Hinaut 2019b]. We also showed that the model could work with grammatical constructions of fifteen different European and Asian lan-

[3] That is why talking about "sensorimotor neurons" instead of "mirror neurons" is more a pragmatic term when considering various species.

[4] In the lab of Henry Kennedy who was working on primate cortical hierarchy.

[5] In short, grammatical constructions are sentence structures in which content words are replaced by "placeholders". The model had to output thematic role labels of these content words. It corresponds to answer the question "Who did what to whom?".

[6] iCub video sentence comprehension: https://www.youtube.com/watch?v=AUbJAupkU4M

[7] iCub video sentence production: https://www.youtube.com/watch?v=3ZePCuvygi0

[8] Nao video: https://www.youtube.com/watch?v=R4cE4bAhLrU.

[9] Switch of languages within the same sentence.

guages [Hinaut & Twiefel 2020]. More recently, we also analysed the ability of reservoirs to generalize with little data in the context of robot language acquisition and compared it to LSTMs [Juven & Hinaut 2020, Dinh & Hinaut 2020, Variengien & Hinaut 2020, Oota *et al.* 2022]. We also made a first version of language processing with *Hierarchical-Task Reservoirs*, from speech to semantic labels [Pedrelli & Hinaut 2020, Pedrelli & Hinaut 2022]. Two of these later works will be presented in Chapter 4.

In parallel, staying on the track of how cortex encodes abstract sequences, I started to work on canaries, which are a good animal model of language acquisition because songbirds and humans share similar vocal developmental stages [Doupe & Kuhl 1999, Pagliarini *et al.* 2021b]. We set up an experimental protocol to study how a sensorimotor area[10] encodes local and global variations of chunks in sequence of syllables in canary songs. We also analysed various syntactic features and visual representations of canary songs [Hinaut *et al.* 2017, Cazala 2019]. These studies included building tools to analyse automatically canary songs [Trouvain & Hinaut 2021] and the release of an open source canary dataset [Giraudon *et al.* 2021].

Furthermore, still in link with songbirds, I started to make models directly in interaction with the acoustic environment, with vocal sensorimotor models – on which we made a review [Pagliarini *et al.* 2021b]. We built a canary sensorimotor model learning to produce syllable with a simple Hebbian rule [Pagliarini *et al.* 2021a]: the motor layer was obtained by constructing a low-dimensional Generative Adversarial Network (GAN) to generate qualitative canary sounds [Pagliarini *et al.* 2021c], and the perceptual layer was performed with a reservoir syllable classifier [Trouvain & Hinaut 2021]. These later works will be presented in Chapter 4.

Finally, we extended the short-term memory of reservoirs by proposing a robust Working Memory (WM) model of prefrontal cortex, able to gate information in line-attractors [Strock *et al.* 2020]; also presented in Chapter 4. An extension of this model to long-term memory was performed with the help of Conceptors [Strock *et al.* 2022]. Besides, we built new tools to analyse reservoir dynamics [Variengien & Hinaut 2020] along with *ReservoirPy* [Trouvain *et al.* 2020, Trouvain *et al.* 2022, Trouvain & Hinaut 2022]; both presented in Chapter 4. ReservoirPy is a flexible Python library which will enable us to quickly build new kinds of reservoir models, from sensorimotor to hierarchical ones. It will hopefully ease the sharing and reuse of reservoir models across the community. In the research program presented in this manuscript I want to catalyse these previous works.

The manuscript is organized as follows. First, I will present my scientific trajectory in Chapter 2 including my CV and list of publications. Then, in Chapter 3, as Reservoir Computing (RC) is central to my research, we will quickly dive in reservoirs in order to give quick intuitions. Afterwards, contributions highlighted in Chapter 2 will be presented in Chapter 4. Thereafter in Chapter 5, I will present

[10] HVC (used as a proper name) area is part of the pallium, bird's equivalent of cortex.

my research program relying on these contributions. Finally in Chapter 6, I will discuss the project, propose some perspectives, along with a thought experiment.

Pluridisciplinarity disclaimer This manuscript will make connections to various scientific fields: I do not pretend to be an expert in all these fields, thus you may expect some approximations at some points. We need to create models that are as simple as possible while having a good explanatory power. Creating models as complex as what we observe would probably not help to understand the core brain mechanisms at play in our everyday behaviors. Pluridisciplinarity is what animates me since my studies in computer science major and cognitive science minor. To understand how the brain works, I believe some of us need to go through this highly interdisciplinary path, taking the risk of several shortcuts in some fields, because borrowing concepts from one science to another, like languages borrow words from one another, is what makes them alive.

Chapter 2

Scientific Trajectory

Contents

2.1 Curriculum Vitae

Personal details	
Gender	M
NAME and first name	HINAUT Xavier
Year of birth	1985
Country	France

Current position

Function

Inria Research Scientist ("Chargé de Recherche – Classe Normale", CR-CN) – (This is a permanent position)

Other activities

Supervision

- (Starting) PhD supervision of Nathan Trouvain (2022-now).
- (Starting) PhD co-supervision of Kalidou Ba (2022-now).
- Current PhD co-supervision of Subba Reddy Oota (2020-now) with Frédéric Alexandre.
- Engineer supervision of Nathan Trouvain (2020-2022).
- PhD co-supervision of Silvia Pagliarini (2017-2021) with Arthur Leblois.
- PhD co-supervision of Anthony Strock (2017-2020) with Nicolas Rougier.
- Post-doc supervision of Luca Pedrelli (2019-2020).
- Overall supervision or co-supervision of 17 Bachelors, 18 Masters, 3 (+2) PhD, 1 Post-Doc and 1 Engineer (since 2014).

Bachelor students: (2014) F. Lavallée; (2015) C. Garber, A. Klassen; (2016) R. Portelas, R. Pastureau, L. Devers, L. Fauvel, R. Confiant-Duté, C. Soetaert, K. Ignatowicz; (2017) Y. Li; (2018) P. Marcus; (2019) J. Giraudon, M. Caute; (2020) A. Variengien; (2021) B. Lhopitallier; (2022) P. Croizet.

Master students: (2014) C Droin; (2015) S Kumar; (2016) E Le Masson; (2017) J-B Zacchello, A Skiada, A Strock; (2019) P Detraz, R Teitgen, A Juven, C. Bezier, C. Chenouna; (2020) T.T. Dinh, N. Trouvain, C. Chauvet; (2021) T. Pemeja (2022) A. Arthuis, T. Barennes, L. Rabastain.

Teaching activities

2020–now: Scientific supervisor student project at the workshop "AI 4 Industry", 1 week every January, Bordeaux, FR.
2016–now: MSc Eng. Bordeaux INP engineering schools, MSc Cognitive Science, University of Bordeaux, FR.
Topics: Modelling, Artificial Neural Networks, Data Mining.
2015–now: Tutoring several Bachelor and Master semester projects.
2015–now: Regular invited lectures to MSc Intelligent Adaptive Systems, University of Hamburg, DE.
2010–2013: MSc Cognitive Science & BSc Computer Science, University of Lyon, FR. & Modelling, Artificial Neural Networks, Applied Mathematics and Programming.

Institutional responsibilities

2022–now: WorkPackage leader "New methods in Machine Learning for Health Data", Public Health Data Science "Impulsion" Network, Bordeaux, FR
2021–now: Chair of the IEEE Task Force Language and Cognition
2021–now: Member of the "Committee for Research Jobs" (CER) of Inria, Bordeaux, FR
Member of the "Committee for Technological Development" of Inria, Bordeaux, FR (a committee selecting technical projects and hiring engineers)
2019–now: Co-Head of the "NeuroRobotics" CNRS Working Group Organisation of several workshops each year
2017–now: Member of the "Committee for Technological Development" of Inria, Bordeaux, FR (a committee dealing with Engineers hiring and engineering projects in the Institute)
2017–now: President of the MindLABx association, Bordeaux, FR. Events on AI and Cognitive Science. Main event in 2017: 3 days hackathon, 50 people, 6 k€, Bordeaux science museum.
2010–2012: Elected member of the "Neuroscience and Cognition" Doctoral School council, PhD student representative, Lyon, FR + Vice-Secretary of FRESCO (national federation of students in cognitive sciences), FR.

Scientific committees & Conference organisation

- **Guest editor**: Frontiers in Robotics and AI: SI on "Language and Robotics" (2021); Advanced Robotics: SI on "Machine Learning Methods for High-Level Cognitive-Capabilities-in-Robotics" (2019)
- Member of the **Editorial Board** of Frontiers in Neurorobotics as Review Editor.
- **Session chair**: ICDL-EpiRob conf, Valparaiso, Chile, Sep 2020. CogSci conference, Montréal, CA, Jul 2019. ICDL-EpiRob conference, Tokyo, JP, Sep 2018.
- **IEEE Task Force (TF) member**: "Reservoir Computing", "Cognitive and Developmental Systems Technical Committee": "Language and Cognition" TF and "Action and Perception" TF.
- **Program committee member of conferences/workshops**: International Conference on Artificial Neural Networks (ICANN) 2021. International Combined Workshop on Spatial Language Understanding and Grounded Communication for Robotics

(SpLU-RoboNLP) 2021. IROS Workshop on Deep Probabilistic Generative Models for Cognitive Architecture in Robotics, Macau, China. Nov 2019. International Workshop on Cognitive and Neural Systems, Granada, Spain. Oct 2019.
- **Workshop/Conference organization**:
2022: Local organizer of Inria-DFKI workshop, Bordeaux, FR (~60 people)
2020-2022 ICDL SMILES workshop (Sensorimotor Interaction, Language and Embodiment of Symbols): 2 online and 1 onsite (at Queen Mary University, London, UK in 2022), from 50 to 120 registrations;
2019: ICDL-EpiRob "Workshop on Language Learning", Oslo, NW;
2018: IROS "Workshop on Language and Robotics", Madrid, ES;
2017: "Workshop on Machine Learning Methods for High-Level Cognitive Capabilities in Robotics", Vancouver, CA;
2016 & 2019: "Day of the NeuroRobotics Working Group", Bordeaux, FR.
- **Reviewer for Journals**: Adaptive Behavior, Applied Science, Cognitive Computation, Cognitive Systems (CogSys), Entropy, Frontiers in Psychology, Front. in Neurorobotics, Intellectica, Neural Networks, PeerJ, PLoS ONE, PLoS Computational Biology, ReScience, ACM Transactions in Human-Robot Interaction (THRI), IIEE Transactions in Neural Networks and Learning Systems (TNNLS), IEEE Transactions in Cognitive Developmental Systems (TCDS).
- **Reviewer for Conferences**: CogSci, ESANN, ICANN, ICDL-EpiRob, IJCNN, SpLU workshop.
- **General public talks or demonstrations:**
"La Course 12--4--90", Art & Science performance at "Drôles d'Objets" conference, La Rochelle, FR, Oct 2021.
"The trial of the robot" virtual interview with high school students, Cap Sciences, Bordeaux. FR.
Invited talk at the Science Fest Day, Cap Sciences, Bordeaux, FR, Oct 2019.
Pint of Science invited talk, Bordeaux, FR, May 2019.
Haillan Library, invited talk, Bordeaux, FR, May 2019.
Demonstration with Nao robot talking, Science fest, Bordeaux, FR, oct 2018.
(+ several other public talks the previous years.)
- General public jury:
Jury evaluator for French-speaking worldwide virtual hackathon "Créathon", Poitiers. May 2019.
- PhD jury member
Florian Golemo in December 2018, at Inria Bordeaux, FR.
- PhD monitoring committee
Gautier Hamon, 2022—now, Inria Bordeaux, FR.
Manel Rakez, 2022—now, Bordeaux University & Inserm, FR.
Tristan Karch, 2019—2022, Inria Bordeaux, FR.
William Schueller, 2018, Inria Bordeaux, FR.

Invited talks to labs and at conferences/workshops
- "Sensorimotor and hierarchical models of vocal and language learning: songbirds, humans, robots." Computational and mathematical approaches for neuroscience workshop, Paris Brain Institute, Paris, FR. June 2022.
- "ReservoirPy: a Python library for Reservoir Computing", Le Lyre lab, Suez, Pessac, FR. May 2021. (online)
- "ReservoirPy: a Python library for Reservoir Computing", Machine Learning in Montpellier (ML-MTP) seminar, FR. May 2021. (online)
- "Building a Vocal Sensorimotor Model for Canaries", European Birdsong Meeting 2022, Capo Caccia, Sardinia, IT. May 2021.
- "Reservoir Computing: de la théorie à la pratique avec ReservoirPy", Engineer Service Inria Paris, SCAI, Sorbonne University, Paris. March 2021.
- "Towards interactive models with Reservoir Computing", HILL seminar, J. Rączaszek-Leonardi lab, University of Warsaw. Dec 2021. (online)
- "Sensorimotor and Hierarchical Models of Vocal and Language Learning: Songbirds, Humans, Robots", S. Franck lab, Nijmegen, NL. Nov 2021.
- "Recurrent Neural Networks" Tutorial, R4 (robotics) workshop, Bidart, FR. Nov 2021.
- "Sensorimotor and Hierarchical Models of Vocal and Language Learning: Songbirds, Humans, Robots", 10th Peripatetic conference Cognitive System Modelling 10th, Zakopane, PL. Oct 2021.
- "Sensorimotor and Hierarchical Modeling of Sequence Processing: Songbirds, Humans, Robots", NeuroFrance, FR. May 2021. (online)
- "How to teach a robot to sing like a bird?", E. Vincent lab, LORIA, Inria, Nancy, FR. March 2021. (online)
- "Reservoir Computing", R4 (robotics) seminar. FR. Feb 2021. (online)
- "Random RNNs to model complex sequence processing: From songbirds to robot learning languages", Cognitive Machine Learning Lab (CoML) Dupoux lab, Paris, FR. May 2020. (online)
- "Une introduction au traitement du langage naturel", AI 4 Industry workshop, Bordeaux, FR. January 2020.
- "How to ground sensorimotor sequence of symbols? From robot learning languages to songbirds", ICDL-EpiRob workshop on Language Learning (M. Spranger), Olso, NW. Aug 2019.
- "Random recurrent networks for language and bird song learning", B. Golosio, Physics dept., University of Cagliary, IT. May 2019.
- "Modélisation de l'encodage neuronal pour l'apprentissage de séquences complexes", LIRIS lab, University of Lyon, FR. June 2019.
- "Modèles Neuronaux Récurrents pour le Traitement de Séquences Complexes", Journée GT-ACAI-Neurorobotique: Neuroscience et multimodalité dans les interactions humain-humain, humain-agent ou humain-robot, Telecom ParisTech, Paris, FR. May 2019.
- "Modelling neuronal encoding of complex sequence learning", "Acoustic Perception" seminar from ETIS lab, University of Cergy-Pontoise, FR. March 2019.
- Symposium organized by regional chair on technological systems for human augmentation, March 28—29, 2019, Bordeaux, France.

- “Neuro-Inspired Model for Robots Learning Language from Phonemes, Words or Grammatical Constructions”, Workshop and Language and Robotics, IROS 2018, Madrid, Spain. Oct 2018.
- "Model of prefrontal cortex and basal ganglia for encoding, learning and producing complex sequences", T.Fukai lab, RIKEN Center for Brain Science, Tokyo, JP. Sept 2018.
- "Model of Prefrontal Cortex for Language Acquisition and Human-Robot Interaction", T. Taniguchi lab, Ritsumeiken University, Kyoto, JP. Sept 2018.
- "Modelling sentence processing with random recurrent neural networks and applications to robotics", Workshop on "The role of the basal ganglia in the interaction between language and other cognitive functions", DEC, Ecole Normale Supérieure (ENS) Ulm, Paris, France. Oct 2017
- "Reservoir Computing for Robot Language Acquisition", 2016 IROS Workshop on Machine Learning Methods for High-Level Cognitive Capabilities in Robotics, Daejon, KR. Oct 2016.
- "Prefrontal cortex model for language acquisition as abstract sequence learning applied to human robot interaction", T. Buschman lab, Princeton Neuroscience Institute (PNI), Princeton University, USA. Dec 2015.
- "Syntax Acquisition with Echo State Networks and application to Human-Robot Interaction", NATS team, University of Hamburg, DE. Nov 2015.
- "How to characterize HVC neuron responses to syntactic properties of the songs?", "Sequence" seminar of Unicog (S. Dehaene’s lab), Neurospin, INSERM-CEA, Saclay. Nov. 2014.
- Invited lecture, in Jean-Louis Dessales’s course, at Télécom Paris, FR. 2014.
- "Recurrent Neural Networks for Grammatical Structure Processing, with an Application to Human-Robot Interaction", [* similar talk in different labs] Neurospin, INSERM-CEA, Saclay, FR. Apr. 2014. & "Sequence" seminar of Unicog (S. Dehaene’s lab), Neurospin, INSERM-CEA, Saclay, FR. Apr. 2014. & ENSTA ParisTech, Flowers lab (PY Oudeyer), Palaiseau, FR. Apr. 2014. & NeuroPSI (C. Del Negro), Orsay, FR. Nov 2013. & Synalp team (C. Gardent & T. Voegtlin), Loria-INRIA, Nancy, FR. Nov 2013. & Knowledge Technology Group (S. Wermter), University of Hamburg, DE. May 2013. & P. Hagoort lab, Max Planck Institute for Psycholinguistics, Nijmegen, NL. Feb 2013.
- "Grammatical Structure Processing and Discussion on ESN Architectures", MINDS lab (H. Jaeger), JACOBS University, Bremen, DE. Nov 2013.

Invited plenary lectures in international master and summer schools

- Invited lecture, LACORO Summer School. Chile (online). Aug 2020.
- Invited lecture, PhD Retreat of Center for Research and Interdisciplinarity (CRI) of Paris. Arcachon 2019
- Invited lecture, Summer school of ETIS lab, University of Cergy-Pontoise. Moliets, FR. Sept 2018.
- Invited lectures on Reservoir Computing and Recurrent Neural Networks, at International MSc "Intelligent Adaptive Systems", University of Hamburg, DE. From 2015 to 2012 (but 2020-2021).

Previous positions

Début / Start date	Fin / End date	Ville / Town	Etablissement / Organisation	Fonction / Function
2015	2016	Hamburg, DE	University of Hamburg	Postdoctoral Marie Curie Fellow
2014	2014	Orsay, FR	CNRS / University of Paris-Sud-Orsay	Postdoctoral Fellow
2013	2013	Hamburg, DE	University of Hamburg	Postdoctoral Fellow
2009	2013	Lyon, FR	Inserm / University of Lyon	PhD student

Career interruption(s)

Education

- 2013: PhD in Computational Neuroscience
  -- Stem Cell and Brain Research Institute, INSERM, Lyon University
  -- PhD Supervisor: Peter F. Dominey
  -- Title: Recurrent Neural Networks for Abstract Sequence and Grammatical Structure Processing, with an Application to Human-Robot Interaction (Awards are no longer given by University of Lyon)
- 2009: École Pratique des Hautes Études (EPHE), Paris, France
  -- Master of Science (MSc), Artificial and Natural Cognition. **Summa cum Laude (Honors)**
- 2008: Université de Technologie de Compiègne (UTC), France
  -- Engineer, Computer Science. Specialty: Data Mining. Minor: Cognitive Science.
  -- MSc, Computer Science. **Magna cum Laude (High honors)**

Productions scientifiques / Scientific productions

Projets de recherche, prix, distinctions, bourses, etc. / Grants, prizes, awards, fellowships, etc.

- 2022-2025: **Young Researcher ANR grant** (French National Agency) “DeepPool”, 302 k€. (as PI)
- 2021: Campus France PHC Van Gogh “BilingualSWITCH”, 1-year travel grant between University of Nijmegen (NL) and Inria, 4k€.
- 2020-2023: 4 year multi-lab ANR “CoBioPro” (biomimetic control of prostheses). (as partner)
- 2020-2023: Inria CORDI-S PhD Fellowship “NewSpeak” project, 111 k€. (as PI)
- 2020-2022: Inria ADT Engineer Fellowship “Scikit-ESN” project for the development of the ReservoirPy library, 80 k€. (as PI)

- 2019-2021: Inria CORDI-S Post-doc Fellowship HURRICANE (deep reservoirs), 65 k€. (as PI)
- 2017-2020: Inria CORDI-S PhD Fellowship “SONNET (songbird sensorimotor model)” project, 100 k€. (as PI)
- 2016-2018: Campus France PHC Procope: “LingoRob” project. 2-year travel grant between University of Hamburg and Inria.
- 2015-2017: **Marie Curie Intra-European Fellowship**. Project: "Echo State Networks for Developing Language Robots". University of Hamburg, Germany. (as a Post-doc)
- 2006-2013: Several Travel Grants for winter and summer schools: Italy, Cambridge, UK, Switzerland, Island.

	5 most relevant publications	What is the major contribution of this publication?
[1]	Strock, A., Hinaut, X.*, Rougier, N.P.* (2020). A Robust Model of Gated Working Memory. Neural computation, 32(1), 153-181.	We were able to model working memory mechanisms by training a simple random recurrent network to gate information for a long period of time. There are four main findings: (1) the model extends the memory capacity of reservoirs with only one continuous output node, (2) the model is very robust against perturbations while not being hand-crafted, (3) we can fit different kinds of dynamics observed in experimental data by simply changing a meta-parameter, (4) we explain the model by deriving a minimal model based on 3 neurons that produce equivalent behavior even in case of lesion. Such “gated reservoir” can be used as a new tool by the reservoir community.
[2]	Pagliarini, S., Leblois, A., Hinaut, X. (2020). Vocal imitation in sensorimotor learning models: a comparative review. IEEE Journal of Transaction in Cognitive Developmental Systems.	This review provides a comparison and synthesis of many computational models of sensorimotor (SM) learning, mostly about vocal learning but not only. Reviewers agreed that such a paper was needed in the community: in fact, computational SM models are difficult to compare because they do not use the same components nor the same architecture.
[3]	Hinaut, X., Dominey, P.F. (2013). Real-Time Parallel Processing of Grammatical Structure in the Fronto- Striatal System: A Recurrent Network Simulation Study Using Reservoir Computing. PloS ONE 8(2): e52946.	This paper presents a sentence comprehension model based on reservoir computing. It provided several novelties compared to previous models: it can generalize to unseen sentence templates (i.e. constructions) on different training corpus sizes, it provides an online prediction of thematic roles during sentence parsing, it provides a hypothesis for the triggering of P600 Event-Related Potential (ERP observed with EEG).
[4]	Pedrelli, L. & Hinaut, X. (2021). Hierarchical-task reservoir for online semantic analysis from continuous speech. IEEE Transactions on Neural Networks and Learning Systems.	We propose a new kind of deep reservoir architecture: the Hierarchical-Task Reservoir (HTR). This architecture proposes to solve a hierarchical task with each reservoir performing a sub-task. We created a challenging task by taking the well-known challenging TIMIT speech dataset: the layers of the reservoirs need to perform online phoneme recognition, word recognition, Part-of-Speech (POS) recognition and Semantic Role Labelling (SRL). We demonstrate that the architecture is able to solve the task with good performance.
[5]	Pagliarini, S., Leblois, A., Hinaut, X. (2021, preprint). What does the Canary Say? Low-Dimensional GAN Applied to Birdsong	We managed to train GANs for raw sounds (WaveGAN) with a large dataset of canary syllables (16000 renditions) and constrain the latent space to small dimensions (from 1 to 6). The sounds produced by the generators were identified and evaluated by a reservoir-based classifier trained on the same dataset, and computer the Inception Score (another quantitative measure). We also performed qualitative evaluation (using UMAP) of the GAN output spectrograms across GAN training epochs and latent dimensions. UMAP representations show the similarities between the training data and the generated data, and between the generated syllables and the interpolations produced. By exploring the latent representations of syllable types, we showed that they form well identifiable subspaces of the latent space.

Valorisation

Founder and manager of ReservoirPy: a flexible Python library for Reservoir Computing.
https://github.com/reservoirpy/reservoirpy
ReservoirPy is a simple user-friendly library based on Python scientific modules. It provides a flexible interface to implement efficient Reservoir Computing (RC) architectures with a particular focus on Echo State Networks (ESN). Advanced features of ReservoirPy allow to improve computation time efficiency on a simple laptop compared to basic Python implementation, with datasets of any size.
Some of its features are: offline and online training, parallel implementation, sparse matrix computation, fast spectral initialization, advanced learning rules (e.g. Intrinsic Plasticity) etc. It also makes possible to easily create complex architectures with multiple reservoirs (e.g. deep reservoirs), readouts, and complex feedback loops. Moreover, graphical tools are included to easily explore hyperparameters with the help of the hyperopt library. It includes several tutorials exploring exotic architectures and examples of scientific papers reproduction. Moreover, graphical tools are included to easily explore hyperparameters with the help of the hyperopt library. It includes a detailed documentation https://reservoirpy.readthedocs.io/ and PyPi package for easy installation.

Useful links

Web page: www.xavierhinaut.com
ORCID: http://orcid.org/0000-0002-1924-1184
Google Scholar: https://scholar.google.com/citations?user=pNW4eZAAAAAJ&hl=fr&oi=ao
GitHub: https://github.com/neuronalX

All my publications are accessible on HAL

Per year: https://tinyurl.com/ydt3z57p (full link below)
http://haltools.inria.fr/Public/afficheRequetePubli.php?auteur_exp=Xavier,Hinaut&CB_auteur=oui&CB_titre=oui&CB_article=oui&langue=Francais&tri_exp=annee_publi&ordre_aff=TA&Fen=Aff&css=../css/VisuOmbreVignettes.css
HAL CV (per type): https://cv.archives-ouvertes.fr/xavier-hinaut

2.2 List of publications

Xavier Hinaut's Bibliography

ORCID: 0000-0002-1924-1184.

LA Language – RB Robotics – RC Reservoir Computing – SB Song Bird – SM SensoriMotor

Preprints

PP.1 S. R. Oota, F. Alexandre, and **X. Hinaut**. "Cross-Situational Learning Towards Robot Grounding". hal-03628290 preprint. Apr. 2022.

PP.2 N. Trouvain and **X. Hinaut**. "reservoirpy: A Simple and Flexible Reservoir Computing Tool in Python". hal-03699931 preprint. June 2022. RC

PP.3 S. Pagliarini, N. Trouvain, A. Leblois, and **X. Hinaut**. "What does the Canary Say? Low-Dimensional GAN Applied to Birdsong". hal-03244723 preprint. Nov. 2021. SB

PP.4 A. Variengien and **X. Hinaut**. "A Journey in ESN and LSTM Visualisations on a Language Task". hal-03030248 preprint. Nov. 2020.

Journals

IJ.1 A. Strock, N. P. Rougier, and **X. Hinaut**. "Latent space exploration and functionalization of a gated working memory model using conceptors". In: *Cognitive Computation* (Jan. 2022). RC

IJ.2 L. Pedrelli and **X. Hinaut**. "Hierarchical-Task Reservoir for Online Semantic Analysis from Continuous Speech". In: *IEEE Transactions on Neural Networks and Learning Systems* (Sept. 2021). LA RC

IJ.3 S. Pagliarini, A. Leblois, and **X. Hinaut**. "Vocal Imitation in Sensorimotor Learning Models: a Comparative Review". In: *IEEE Transactions on Cognitive and Developmental Systems* (Nov. 2020). SB SM

IJ.4 **X. Hinaut** and J. Twiefel. "Teach Your Robot Your Language! Trainable Neural Parser for Modelling Human Sentence Processing: Examples for 15 Languages". In: *IEEE Transactions on Cognitive and Developmental Systems* (Dec. 2019). LA RB RC

IJ.5 A. Strock, **X. Hinaut**, and N. P. Rougier. "A Robust Model of Gated Working Memory". In: *Neural Computation* (Nov. 2019), pp. 1–29. RC

IJ.6 N. P. Rougier, K. Hinsen, F. Alexandre, T. Arildsen, L. Barba, F. C. Y. Benureau, C. T. Brown, P. de Buyl, O. Caglayan, A. P. Davison, M. A. Delsuc, G. Detorakis, A. K. Diem, D. Drix, P. Enel, B. Girard, O. Guest, M. G. Hall, R. N. Henriques, **X. Hinaut**, K. S. Jaron, M. Khamassi, A. Klein, T. Manninen, P. Marchesi, D. Mcglinn, C. Metzner, O. L. Petchey, H. E. Plesser, T. Poisot, K. Ram, Y. Ram, E. Roesch, C. Rossant, V. Rostami, A. Shifman, J. Stachelek, M. Stimberg, F. Stollmeier, F. Vaggi, G. Viejo, J. Vitay, A. Vostinar, R. Yurchak, and T. Zito. "Sustainable computational science: the ReScience initiative". In: *PeerJ Computer Science* 3 (Dec. 2017). 8 pages, 1 figure, e142.

IJ.7 **X. Hinaut**, F. Lance, C. Droin, M. Petit, G. Pointeau, and P. Dominey. "Corticostriatal response selection in sentence production: Insights from neural network simulation with reservoir computing". In: *Brain and Language* 150 (Nov. 2015), pp. 54–68. LA RC

IJ.8 **X. Hinaut**, M. Petit, G. Pointeau, and P. F. Dominey. "Exploring the acquisition and production of grammatical constructions through human-robot interaction with echo state networks". In: *Frontiers in Neurorobotics* 8 (May 2014). LA RB RC

IJ.9 **X. Hinaut** and P. F. Dominey. "Real-Time Parallel Processing of Grammatical Structure in the Fronto-Striatal System: A Recurrent Network Simulation Study Using Reservoir Computing". In: *PLoS ONE* 8.2 (Feb. 2013), e52946. LA RC

IJ.10 **X. Hinaut** and P. F. Dominey. "A three-layered model of primate prefrontal cortex encodes identity and abstract categorical structure of behavioral sequences". In: *Journal of Physiology - Paris* 105.1-3 (Jan. 2011), pp. 16–24. RC

Conferences

IC.1 N. Trouvain, N. P. Rougier, and **X. Hinaut**. "Create Efficient and Complex Reservoir Computing Architectures with ReservoirPy". In: *SAB 2022 - FROM ANIMALS TO ANIMATS 16: The 16th International Conference on the Simulation of Adaptive Behavior*. Cergy-Pontoise / Hybrid, France, Sept. 2022. RC

IC.2 **X. Hinaut**. "La Course 12–4–90". In: *Drôles d'objets 2021 - Un nouvel art de faire*. La Rochelle, France, Oct. 2021. RB

IC.3 **X. Hinaut** and N. Trouvain. "Which Hype for my New Task? Hints and Random Search for Reservoir Computing Hyperparameters". In: *ICANN 2021 - 30th International Conference on Artificial Neural Networks*. Bratislava, Slovakia, Sept. 2021. RC

IC.4 S. Pagliarini, A. Leblois, and **X. Hinaut**. "Canary Vocal Sensorimotor Model with RNN Decoder and Low-dimensional GAN Generator". In: *ICDL 2021- IEEE International Conference on Development and Learning*. Beijing, China, Aug. 2021. RC SB SM

IC.5 N. Trouvain and **X. Hinaut**. "Canary Song Decoder: Transduction and Implicit Segmentation with ESNs and LTSMs". In: *ICANN 2021 - 30th International Conference on Artificial Neural Networks*. Vol. 12895. Farkaš I., Masulli P., Otte S., Wermter S. (eds) Artificial Neural Networks and Machine Learning – ICANN 2021. Lecture Notes in Computer Science. Bratislava, Slovakia: Springer, Cham, Sept. 2021, pp. 71–82. RC SB

IC.6 T. T. Dinh and **X. Hinaut**. "Language Acquisition with Echo State Networks: Towards Unsupervised Learning". In: *ICDL 2020 - IEEE International Conference on Development and Learning*. Valparaiso / Virtual, Chile, Oct. 2020. LA RB RC

IC.7 A. Juven and **X. Hinaut**. "Cross-Situational Learning with Reservoir Computing for Language Acquisition Modelling". In: *2020 International Joint Conference on Neural Networks (IJCNN 2020)*. Glasgow, Scotland, United Kingdom, July 2020. LA RB RC

IC.8 L. Pedrelli and **X. Hinaut**. "Hierarchical-Task Reservoir for Anytime POS Tagging from Continuous Speech". In: *2020 International Joint Conference on Neural Networks (IJCNN 2020)*. Glasgow, Scotland, United Kingdom, July 2020. LA RC

IC.9 N. Trouvain, L. Pedrelli, T. T. Dinh, and **X. Hinaut**. "ReservoirPy: an Efficient and User-Friendly Library to Design Echo State Networks". In: *ICANN 2020 - 29th International Conference on Artificial Neural Networks*. Bratislava, Slovakia, Sept. 2020. RC

IC.10 **X. Hinaut** and M. Spranger. "Learning to Parse Grounded Language using Reservoir Computing". In: *ICDL-Epirob 2019 - Joint IEEE 9th International Conference on Development and Learning and Epigenetic Robotics*. Olso, Norway, Aug. 2019. LA RB RC

IC.11 A. Strock, N. P. Rougier, and **X. Hinaut**. "Using Conceptors to Transfer Between Long-Term and Short-Term Memory". In: *Artificial Neural Networks and Machine Learning (ICANN) 2019*. Munich, Germany: Springer, Sept. 2019, pp. 19–23. RC

IC.12 S. Pagliarini, **X. Hinaut**, and A. Leblois. "A bio-inspired model towards vocal gesture learning in songbird". In: *2018 Joint IEEE International Conference on Development and Learning and Epigenetic Robotics (ICDL-EpiRob)*. Corresponding code at https://github.com/spagliarini/2018-ICDL-EPIROB. Tokyo, Japan, Sept. 2018. SB SM

IC.13 S. Pagliarini, **X. Hinaut**, and A. Leblois. "Learning an inverse model for vocal production: toward a bio-inspired model". In: *European Birdsong Meeting*. Odense, Denmark, Apr. 2018. SB SM

IC.14 A. Strock, N. P. Rougier, and **X. Hinaut**. "A Simple Reservoir Model of Working Memory with Real Values". In: *2018 International Joint Conference on Neural Networks (IJCNN)*. Rio de Janeiro, Brazil, July 2018. RC

IC.15 J. Twiefel, **X. Hinaut**, and S. Wermter. "Syntactic Reanalysis in Language Models for Speech Recognition". In: *2017 Joint IEEE International Conference on Development and Learning and Epigenetic Robotics (ICDL-EpiRob)*. Lisbon, Portugal, Sept. 2017. LA

IC.16 L. Mici, **X. Hinaut**, and S. Wermter. "Activity recognition with echo state networks using 3D body joints and objects category". In: *European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN)*. Bruges, Belgium, Apr. 2016, pp. 465–470. RC

IC.17 J. Twiefel, **X. Hinaut**, M. Borghetti, E. Strahl, and S. Wermter. "Using Natural Language Feedback in a Neuro-inspired Integrated Multimodal Robotic Architecture". In: *25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. Proceedings of the 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN). New York City, United States, Aug. 2016, pp. 52–57. LA RB

IC.18 J. Twiefel, **X. Hinaut**, and S. Wermter. "Semantic Role Labelling for Robot Instructions using Echo State Networks". In: *European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN)*. Bruges, Belgium, Apr. 2016. LA RB RC

IC.19 **X. Hinaut**, J. Twiefel, M. Petit, P. Dominey, and S. Wermter. "A Recurrent Neural Network for Multiple Language Acquisition: Starting with English and French". In: *Proceedings of the NIPS Workshop on Cognitive Computation: Integrating Neural and Symbolic Approaches (CoCo 2015)*. Montreal, Canada, Dec. 2015. LA RC

IC.20 **X. Hinaut** and S. Wermter. "An Incremental Approach to Language Acquisition: Thematic Role Assignment with Echo State Networks". In: *In: Wermter S. et al. (eds) Artificial Neural Networks and Machine Learning – ICANN 2014*. Ed. by W. S. et al. Vol. 8681. Lecture Notes in Computer Science. Hamburg, Germany, Sept. 2014, pp. 33–40. LA RC

IC.21 **X. Hinaut** and P. F. Dominey. "On-Line Processing of Grammatical Structure Using Reservoir Computing". In: *In A. E. P. Villa, et al.: Artificial Neural Networks and Machine Learning - ICANN 2012 - 22nd International Conference on Artificial Neural Networks*. Vol. 7553. Lecture Notes in Computer Science. Lausanne, Switzerland, Sept. 2012, pp. 596–603. LA RC

IC.22 **X. Hinaut** and P. Dominey. "A three-layered cortical network encodes identity and abstract categorical structure of behavioral sequences as in the primate lateral prefrontal cortex". In: *Cinquième conférence plénière française de Neurosciences Computationnelles, "Neurocomp'10"*. Lyon, France, Aug. 2010. RC

Short Papers/Posters

PO.1 **X. Hinaut** and N. Trouvain. "ReservoirPy: Efficient Training of Recurrent Neural Networks for Time-series Processing". In: *EuroSciPy 2022 - 14th European Conference on Python in Science*. Poster. Aug. 2022. RC

PO.2 S. Reddy Oota, F. Alexandre, and **X. Hinaut**. "Long Short-Term Memory of Language Models for Predicting Brain Activation During Listening to Stories". In: *CogSci 2022 - Cognitive Science Society*. Toronto, Canada, July 2022. LA RC

PO.3 S. Reddy Oota, F. Alexandre, and **X. Hinaut**. *Investigating Long-Term Context of Language Models on Brain Activity during Narrative Listening in fMRI*. Poster. July 2022. LA RC

PO.4 S. Reddy Oota, F. Alexandre, and **X. Hinaut**. "Learning to Parse Sentences with Cross-Situational Learning using Different Word Embeddings Towards Robot Grounding". In: *Spatial Language Understanding and Grounded Communication for Robotics Workshop, ACL-IJCNLP 2021*. Aug. 2021. LA RB RC

PO.5 **X. Hinaut** and A. Variengien. *Recurrent Neural Networks Models for Developmental Language Acquisition: Reservoirs Outperform LSTMs*. Poster. Oct. 2020. LA RC

PO.6 P. Detraz and **X. Hinaut**. "A Reservoir Model for Intra-Sentential Code-Switching Comprehension in French and English". In: *CogSci'19 - 41st Annual Meeting of the Cognitive Science Society*. Montréal, Canada, July 2019. LA RC

PO.7 **X. Hinaut**. *From Phonemes to Sentence Comprehension: A Neurocomputational Model of Sentence Processing for Robots*. Poster. May 2018. LA RB RC

PO.8 **X. Hinaut**. "Which Input Abstraction is Better for a Robot Syntax Acquisition Model? Phonemes, Words or Grammatical Constructions?" In: *2018 Joint IEEE International Conference on Development and Learning and Epigenetic Robotics (ICDL-EpiRob)*. Tokyo, Japan, Sept. 2018. LA RB RC

PO.9 S. Pagliarini, A. Leblois, and **X. Hinaut**. "Towards Biological Plausibility of Vocal Learning Models: a Short Review". In: *ICDL-Epirob Workshop on Continual Unsupervised Sensorimotor Learning*. Tokyo, Japan, Sept. 2018. LA SB SM

PO.10 A. Strock, N. P. Rougier, and **X. Hinaut**. "A Simple Reservoir Model of Working Memory with Real Values". In: *Third workshop on advanced methods in theoretical neuroscience*. June 2018. RC

PO.11 J.-B. Zacchello, **X. Hinaut**, and A. Leblois. "Replication of Laje & Mindlin's model producing synthetic syllables". In: *European Birdsong Meeting*. Poster. Apr. 2018. SB

PO.12 **X. Hinaut**. "From Phonemes to Robot Commands with a Neural Parser". In: *IEEE ICDL-EPIROB Workshop on Language Learning*. Lisbon, Portugal, Sept. 2017, pp. 1–2. LA RB RC

PO.13 **X. Hinaut**, A. Cazala, and C. del Negro. "Neural coding of variable song structure in the songbird". In: *EBM 2017 - European Birdsong Meeting*. Bordeaux, France, May 2017, p. 1. SB

PO.14 **X. Hinaut**. "Recurrent Neural Network for Syntax Learning with Flexible Representations". In: *IEEE ICDL-EPIROB Workshop on Language Learning*. Cergy, France, Dec. 2016. LA RC

PO.15 **X. Hinaut** and J. Twiefel. "Recurrent Neural Network Sentence Parser for Multiple Languages with Flexible Meaning Representations for Home Scenarios". In: *IROS Workshop on Bio-inspired Social Robot Learning in Home Scenarios*. Daejon, South Korea, Oct. 2016. LA RB RC

PO.16 **X. Hinaut**, J. Twiefel, and S. Wermter. "Recurrent Neural Network for Syntax Learning with Flexible Predicates for Robotic Architectures". In: *The Sixth Joint IEEE International Conference Developmental Learning and Epigenetic Robotics (ICDL-EPIROB)*. Cergy, France, Sept. 2016. LA RB RC

PO.17 G. Pointeau, M. Petit, **X. Hinaut**, G. Gibert, and P. F. Dominey. "On-Line Learning of Lexical Items and Grammatical Constructions via Speech, Gaze and Action-Based Human-Robot Interaction". In: *INTERSPEECH 2013 - 14th Annual Conference of the International Speech Communication Association*. Lyon, France, Aug. 2013. LA RB RC

PO.18 **X. Hinaut**, M. Petit, and P. Dominey. "Online Language Learning to Perform and Describe Actions for Human-Robot Interaction". In: *Post-Graduate Conference on Robotics and Development of Cognition*. Lausanne, Switzerland, Sept. 2012. LA RB RC

Correspondences

CO.1 F. Alexandre, **X. Hinaut**, N. P. Rougier, and T. Viéville. "Higher Cognitive Functions in Bio-Inspired Artificial Intelligence". In: *ERCIM News*. Special topic "Brain inspired computing" 125 (Apr. 2021).

CO.2 T. Taniguchi, T. Horii, **X. Hinaut**, M. Spranger, D. Mochihashi, and T. Nagai. "Editorial: Language and Robotics". In: *Frontiers in Robotics and AI* 8 (Apr. 2021).

CO.3 T. Inamura, H. Yokoyama, E. Ugur, **X. Hinaut**, M. Beetze, and T. Taniguchi. "Section focused on machine learning methods for high-level cognitive capabilities in robotics". In: *Advanced Robotics* 33.11 (June 2019), pp. 537–538. LA RB

Datasets

PO.1 J. Giraudon, N. Trouvain, A. Cazala, C. Del Negro, and **X. Hinaut**. *Labeled songs of domestic canary M1-2016-spring (Serinus canaria)*. 2021. SB

Technical Reports

TR.1 T. Viéville, **X. Hinaut**, T. F. Drumond, and F. Alexandre. *Recurrent neural network weight estimation through backward tuning*. Research Report RR-9100. Inria Bordeaux Sud-Ouest, Oct. 2017, pp. 1–54.

Book & Thesis

BK.1 **X. Hinaut**. "Recurrent Neural Networks for Abstract Sequence and Grammatical Structure Processing, with an Application to Human-Robot Interaction". PhD thesis. University of Lyon I, France, 2013. LA RB RC

Videos

SO.1 **X. Hinaut**, J. Twiefel, M. B. Soares, P. Barros, L. Mici, and S. Wermter. "Humanoidly Speaking -How the Nao humanoid robot can learn the name of objects and interact with them through common speech". In: *International Joint Conference on Artificial Neural Networks – IJCAI, Video Competition*. Buenos Aires, Argentina, July 2015. LA RB RC

Science Outreach

SO.1 F. Alexandre, D. Chiron, I. Chraibi Kaadoud, M. Courbin-Coulaud, S. Dagar, T. Firmo-Drumond, C. Héricé, **X. Hinaut**, B. T. Nallapu, B. Ninassi, G. Padiolleau, S. Pagliarini, S. de Quatrebarbes, N. P. Rougier, R. Sankar, A. Strock, and T. Viéville. *Neurosmart, une histoire de cerveau et de passionné×e×s de science*. Technical Report RT-0509. Inria, Nov. 2020, p. 19.

Invited Talks

IV.1 **X. Hinaut**. "Modelling sentence processing with random recurrent neural networks and applications to robotics". In: *Workshop "The role of the basal ganglia in the interaction between language and other cognitive functions"*. Anne-Catherine Bachoud-Lévi, Maria Giavazzi, Charlotte Jacquemot, Laboratoire de NeuroPsychologie Interventionnelle. Paris, France, Oct. 2017. LA RB RC

IV.2 **X. Hinaut**. "Reservoir Computing for Robot Language Acquisition". In: *IROS Workshop on Machine Learning Methods for High-Level Cognitive Capabilities in Robotics*. Daejon, South Korea, Oct. 2016. LA RB RC

Tutorials

TU.1 N. Trouvain and **X. Hinaut**. "Reservoir Computing : théorie, intuitions et applications avec ReservoirPy". In: *Plate-Forme Intelligence Artificielle (PFIA)*. Bordeaux, France, June 2021. RC

Chapter 3

Diving into Reservoirs

Contents

Before going more in depth in my research, I will briefly introduce the Reservoir Computing (RC) paradigm. It is central in my work since the beginning of my PhD thesis, during which I worked within the FP7 European Project *Organic* which gathered the European founders of Reservoir Computing. That's why I want to make a short overview to enable readers to have a little idea of what is RC before what will follow.

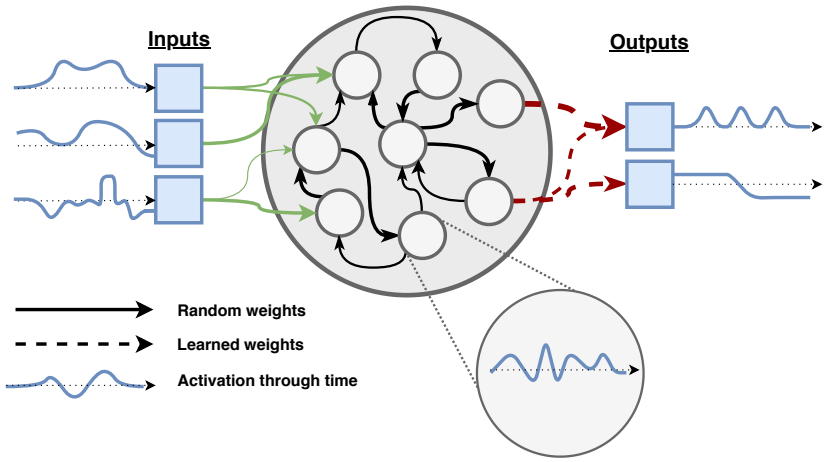

Figure 3.1: The Reservoir Computing (RC) paradigm to train Recurrent Neural Networks (RNNs). Input and recurrent weights are fixed and random while output weights are trained. Time series provided as input generate a non-linear combination of dynamics inside the *reservoir* – the recurrent part in the middle. The output layer linearly *reads out* some of these dynamical combinations – it makes a weighted sum of reservoir states. Image from [Juven & Hinaut 2020].

Reservoir insight. To start, let's dive in reservoir computing with a quick example in Figure 3.1. Inputs are fed to a recurrent layer of neurons, called the *reservoir*.

Reservoir states combine these incoming inputs together with its previous states thanks to the recurrent connections. The reservoir states are also sent to an output layer, called the *read-out*. Input connections and recurrent connections are often fixed and random. Usually, only the output layer connections are trained in a supervised way with a variant of linear regression. We will later see in more details how it works exactly. First, let's jump to the context in which it appeared.

3.1 Context

Reservoir Computing emerged several times. It is often stated that RC has emerged twice in 1995 [Buonomano & Merzenich 1995, Dominey 1995] from the computational neuroscience side, although it can be argued that similar forms have appeared previously several times (see the references collected by Herbert Jaeger on Scholarpedia[1] [Jaeger 2007]). Thus, it appeared only some years after the famous Simple Recurrent Network (SRN) from Elmann in 1990, which was itself featured a few years after the invention of Back-Propagation Through Time (BPTT) [Werbos 1988, Werbos 1990]. Thus, Reservoir Computing can be seen as a possible "end of the road" of simplification of Recurrent Neural Network (RNN) training: first RNNs were fully trained with BPTT, then only one step back in time of BPTT is performed with SRNs, and finally inputs and recurrent weights are not learnt anymore with the RC paradigm.

RC has again emerged in early 2000's with the Echo State Network (ESN) of Jaeger [Jaeger 2001] and with the Liquid State Machines (LSM) of Wolfgang Maass and colleagues [Maass *et al.* 2002]. A RC community started to take shape: machine learning community was more focused on ESNs and computational neuroscience more on LSMs [2]. This movement was probably enhanced because of the nice performances obtained by Jaeger on chaotic time series prediction [Jaeger & Haas 2004].

Some authors did go further in trying to "simplify" the reservoir by removing as much randomness as possible [Rodan & Tino 2010]. In my opinion, randomness seems one of the simplest and most efficient way one can get from a biological point of view, at least to obtain generic computational properties (see Biology paragraph). Creating random neuronal networks seems simple: it does not require to have specific gene expression or other regulatory process for controlling precisely the connections. On the other side of the spectrum, Long Short-Term Memory network (LSTM) coined in 1997 [Hochreiter & Schmidhuber 1997] were another answer to the "Hard Problem" that we will describe now.

Hard problem. Indeed, training connections of a RNN with classical back-propagation through time is known to be a hard problem [Bengio *et al.* 1994, Pascanu *et al.* 2013]. Because, the error gradient tends to vanish or explode when going further back in time in order to capture longer time dependencies. Intuitively,

[1] http://www.scholarpedia.org/article/Echo_state_network

[2] Even if ESNs or equivalent (rate-coded RNNs) were also used in computational neuroscience, e.g.

changing one connection of one neuron can have an impact on all neurons a few timesteps later. That's why Back-Propagation have to be applied "Through Time" (BPTT), in order to send the error gradient "back in time", like a time machine that will change the past in order to change the "present" error. This is done by taking care of the unrolling of events in between[3]. It is hard, because this time machine can "loose track" of the changes needed while going back in time: the error gradient either decreases so much that no connections are modified anymore, or increases so much that the modifications become exponentially huge. In both cases this means that no learning can occur anymore far enough in time.

The LSTM network [Hochreiter & Schmidhuber 1997] was created in order to solve this problem of vanishing or exploding gradient. LSTMs have internal recurrent units that were engineered to enable the BPTT algorithm to be more effective by enabling the error gradient to be kept constant. Inside each unit, they have three (for the 1997 original version [Hochreiter & Schmidhuber 1997]) or four (with an additional forget gate [Gers *et al.* 2000]) parameters. Input gate, forget gate, output gate and the "cell" state. This cell state is the special one that enables to keep the gradient constant if needed: this is the solution provided by LSTMs in order to prevent the gradient from vanishing or exploding. Even though it is an elegant solution, it makes the LSTMs more demanding to train in computational resources because it has more parameters. That's why LSTMs became very popular only in the 2010's with the revolution of deep learning due to new mathematical and implementation tricks [Martens *et al.* 2010, Martens & Sutskever 2011, Sutskever *et al.* 2011, Pascanu *et al.* 2013] along with the popularization of Graphical Processing Units (GPUs) enabling to train these networks faster.

Biology. As we said earlier, Reservoir Computing (RC) emerged at start from the computational neuroscience side [Buonomano & Merzenich 1995, Dominey 1995, Dominey *et al.* 1995, Maass *et al.* 2002], before emerging also in the machine learning side [Jaeger 2001, Jaeger 2002, Jaeger & Haas 2004]. Indeed, a reservoir can be seen as "a canonical computation unit" [Haeusler & Maass 2007]; it could model "a cortical column": what computational neuroscientists often consider as a generic unit of computation. Since 1995 [Dominey 1995], my PhD supervisor Peter Dominey have used it to model the cortico-basal network: the reservoir playing the role of the (prefrontal) cortex and the output layer playing the role of the striatum (input of the basal ganglia from the cortex). Dominey [Dominey *et al.* 1995] showed that even with random networks (that were not called reservoirs yet) it was possible to observe similar neuronal activation patterns then in studies on sequence processing in monkey prefrontal cortex [Barone & Joseph 1989]. RC developed much faster in the machine learning community since the 2000's, but in the 2010's it became more popular from the experimental neuroscientists side. Neuroscientists started using this idea of high-dimensional non-linear representations that can be decoded by a linear classifier. It was a new way to interpret electrophysiological recordings from monkeys [Machens *et al.* 2010, Rigotti *et al.* 2013, Enel *et al.* 2016]: the

[3] Which is not usually the case for time machines.

idea was no longer to find particular sequential pattern in neural activity (like in [Barone & Joseph 1989]), but rather to just decode linearly if some information were present.

3.2 Intuitions in (almost) one page

Short definition: Reservoir Computing is a paradigm to train Recurrent Neural Networks (RNN) without training all connections.

Intuition. The names "reservoir" for the recurrent layer, and "read-out" for the output layer, come from the fact that a lot of input combinations are made inside the recurrent layer (thanks to random projections). The "reservoir" is literally a reservoir of calculations (= "reservoir computing") that are non-linear. From this "reservoir" one linearly decodes (= "reads-out") the combinations that will be useful for the task to be solved. Reservoirs can be implemented on various kinds of physical substrates [Tanaka *et al.* 2019] (e.g. electronic, photonic, mechanical RC).

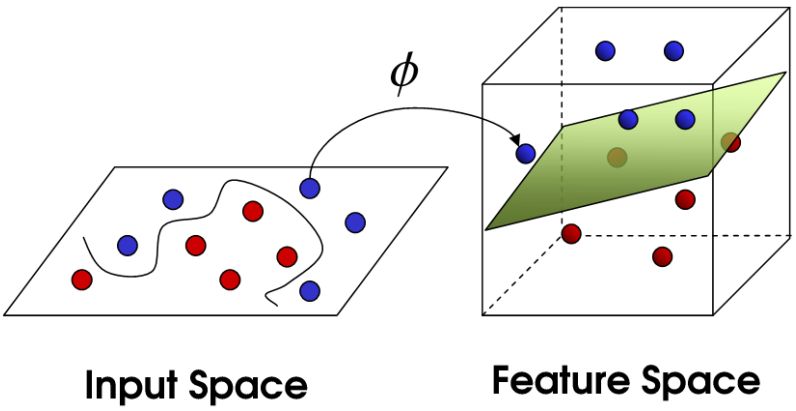

Figure 3.2: Projection of inputs in a higher dimensional space.

The kernel trick. An intuitive way to understand how reservoir computing works is to think it as a temporal Support Vector Machine (SVM) [Verstraeten 2009]. Like in Figure 3.2, suppose you want to separate blue dots from red dots, but in your initial 2D space you cannot separate them with a line. With a SVM [Vapnik 1999] you project theses inputs (i.e. the dots) into a higher dimensional space. In this high dimensional space you can find a hyperplane (an equivalent of a *line* in higher dimensions) that separates your blue dots from your red dots. Finding this hyperplane is equivalent to perform a *linear* regression. You can have different types of kernel with an SVM; in reservoirs this kernel is random.

Multi-task hub. Once a reservoir is trained for a task, it can still be used for another task. Since the computations inside the reservoir are independent of the read-out layer (if there is no feedback connections), new read-out units can be connected to perform a new task. Thus, a reservoir can be seen as a "multi-hub task". This property is interesting to understand how brain areas could share computations: some areas compute and represent information in a way that could be used by several other areas. This "hub" area do not have to compute anything specific or represent information useful for one particular "task": it just have to make "some kind of non-linear computation". The "useful information" is only computed when *reading-out* and projecting to another area. As we discussed in the Introduction Chapter 1, Broca area (LIFG) seems to be involved in representing hierarchical-like structures for language, sequence of actions, music, etc.

Less training data? If we come back to the idea that reservoir computing is like having a temporal SVM, we can imagine that we do not necessarily need much data points to be able to draw a hyperplane to separate our data. Indeed, a SVM can only keep track of points that are close to the the decision boundary[4] – the support vectors. In practice, we have shown that reservoirs needed less data to generalize on an audio classification task [Trouvain & Hinaut 2021] and a on a language task [Variengien & Hinaut 2020, Oota *et al.* 2022].

Extended definition: Reservoir Computing is a paradigm that can use any physical substrate to obtain a suitable combination of inputs before using a read-out layer to extract information from this representational layer (to predict, classify, generate, ...).

We see now that this "reservoir of computation" do not have to be fixed, it can change and adapt over time, for example with homeostatic rules. More importantly it does not need to be computer-based.

3.3 Some equations

There can be different kinds of units in a reservoir: spiking or non-spiking (average firing rate) neurons. There are different kinds of equations for both. I will not speak about spiking version of reservoirs, because I am less familiar with their dynamics[5].

One of the general ways to define ESN is as follows. The state transition of the ESN is computed as follows:

$$x(t) = (1-\alpha)x(t-1) + \alpha\,\mathbf{tanh}(W_{in}u(t) + Wx(t-1)) \tag{3.1}$$

where $u(t) \in R^{N_U}$ is the input vector at time t, $x(t) \in R^{N_R}$ is the reservoir state, $W_{in} \in R^{N_R \times N_U}$ is the input matrix, $W \in R^{N_R \times N_R}$ is the recurrent matrix, $\alpha \in [0, 1]$

[4] This is particularly useful for online version of SVMs to save memory and computation time.

[5] However, I look forward to compare dynamics of spiking and rate-coding neurons as we plan to include spiking neurons inside ReservoirPy (see Chapter 4 Subsection 4.4).

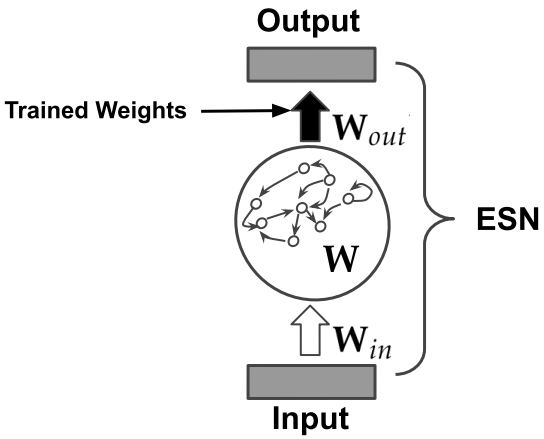

Figure 3.3: **An example of Echo State Network (ESN) architecture (without feedback). Image from [Pedrelli & Hinaut 2020].**

is the leaking rate – more often called the *leak-rate* – and **tanh** is the element-wise hyperbolic tangent. N_U is the number of input units and N_R the number of units in the reservoir. The leak-rate is equivalent to the inverse of a time constant, it is a simplification of writing:

$$\alpha = \frac{dt}{\tau} \tag{3.2}$$

with τ the time constant of neurons and dt the time step discretisation (which equals 1 by default)[6].

The values of matrix W are randomly initialized, for instance using a uniform distribution and then rescaled. This rescaling of W is done in order to obtain a spectral radius[7] ρ equal to the one set by the user as hyperparameter (HP)[8]. The values in matrix W_{in} are randomly initialized, for instance from a uniform distribution and then rescaled in order to have an *input scaling* of σ, which is the one set by the user as hyperparameter. Usually, W and W_{in} matrices are sparse: my recommendation is to use a percentage of non-zero connection of about $10 - 20\%$, but the influence of the sparseness on the performance is often weak. A sparse reservoir enables faster computations.

The output of the ESN is computed as follows:

$$y(t) = W_{out}[1; x(t)] \tag{3.3}$$

where $y(t) \in R^{N_Y}$ is the output at time t, W_{out} is the output matrix, and $[.;.]$ stands for the concatenation of two vectors. N_Y is the number of output (read-out) units.

[6] We showed in [Hinaut & Dominey 2013] that changing dt does not affect much the performance on a language task as soon as the sampling rate of inputs are changed accordingly.

[7] The spectral radius is the maximum absolute eigenvalue of the matrix W.

[8] A hyperparameter is a parameter that need to be predefined and which is not optimized by the learning algorithm.

The output weights are learned using an equivalent of linear regression. The most common practice is to use a regularized version, like the ridge regression:

$$W_{\text{out}} = YX^T(XX^T + \beta I)^{-1} \tag{3.4}$$

where X is the concatenation of the reservoir activities at all time steps with a bias vector at 1, each row corresponding to a time step. Similarly, Y is the concatenation of desired outputs and β is the regularization parameter (often called *ridge* parameter).

A few more details. The spectral radius ρ controls the internal dynamics: more stable dynamics will be obtained for low values and more chaotic ones with high values. I will not talk about the Echo State Property (ESP) as it is a theoretical recommendation from Jaeger [Jaeger 2001] (derived from principles of linear networks) but not a rule that should be followed blindly. In practice spectral radii higher than one should be always tried when exploring hyperparameters because an ESN is a non-linear system that depends on its inputs. Especially, in the case of the leaky ESN where the *effective spectral radius* is different from the one defined by the user [Jaeger *et al.* 2007][9].

You can find a tutorial to explore the hyperparameters of reservoirs in the GitHub repository of our *ReservoirPy* library[10]. We illustrate plots to show how the internal dynamics of the network change with respect to the changes of hyperparameters such as the spectral radius, the input scaling and the leak-rate.

In most our studies, we are use ESNs as defined by Jaeger[11] [Jaeger 2001, Jaeger *et al.* 2007], where the state of each unit also corresponds to its output (i.e. the activation function applies to the states directly). One may argue that it is less biologically plausible, but it has the advantage of having bounded states which prevents the states to take infinite values – which would stop the program because *Not A Number (NAN)* values are encountered. Of course bounded states are obtained with a bounded activation function: e.g. *hyperbolic tangent (tanh)*. This is one of the reasons why we use Jaeger's definition of ESNs. To my knowledge, they seem to be the most used type of reservoir since two decades. Another reason is that it enables to compare our models with many other published papers. For a detailed explanation of the various version of ESNs, David Verstraeten provides a clear explanation in his PhD thesis [Verstraeten 2009].

[9] I have unpublished results showing that one can have very high values of spectral radius (e.g. a million) that still work for a given task as soon as one also decrease the leak-rate. Hyperparameters such as the spectral radius, the leak-rate and the input scaling are linked, that is why we suggest to fix at least one of them when doing hyperparameter exploration [Hinaut & Trouvain 2021].

[10] `https://github.com/reservoirpy/reservoirpy/blob/master/tutorials/4-Understand_and_optimize_hyperparameters.ipynb`

[11] In particular we often use the "leaky" version of ESNs.

Chapter 4

Selected Contributions

Contents

I chose to select eight papers that reflect different pluridisciplinary parts of my work in relation to my Research Program. In order to keep this manuscript not too long, easy to browse and light to download, I made the choice to put only the references to the papers by providing open access links, abstracts and links to figures of the Research Program. Parts of these selected papers are summarized in the context of my Research Program in Chapter 5 in Section 5.1.3.

4.1 Language processing

4.1.1 Towards language acquisition for robots

4.1.1.1 Context

This study [Oota *et al.* 2022] was performed during the PhD thesis of Subba Oota (2020–now) co-supervised with Frédéric Alexandre. It followed the works started during the MSc. internships of Alexis Juven [Juven & Hinaut 2020] (spring 2019) and Trung Dinh [Dinh & Hinaut 2020] (spring 2020), and the BSc. internship of Alexandre Variengien [Variengien & Hinaut 2020] (spring 2020). A summary of the

experiment is proposed in Figure 5.5 of Chapter 5.

Subba Reddy Oota, Frédéric Alexandre, Xavier Hinaut (2022) Cross-Situational Learning Towards Robot Grounding. HAL preprint hal-03628290.

- Open access / HAL version:
  https://hal.archives-ouvertes.fr/hal-03628290
- Supplementary data:
  Directly available in the preprint.

4.1.1.2 Abstract

How do children acquire language through unsupervised or noisy supervision? How does their brain process language? We take this perspective to machine learning and robotics, where part of the problem is understanding how language models can perform grounded language acquisition through noisy supervision and discussing how they can account for brain learning dynamics. Most prior works have tracked the co-occurrence between single words and referents to model how infants learn word-referent mappings. This paper studies Cross-Situational Learning (CSL) with full sentences: we want to understand brain mechanisms that enable children to learn mappings between words and their meanings from full sentences in early language learning. We investigate the CSL task on a few training examples with two sequence-based models: (i) Echo State Networks (ESN) and (ii) Long-Short Term Memory Networks (LSTM). Most importantly, we explore several word representations including One-Hot, GloVe, pretrained BERT, and fine-tuned BERT representations (last layer token representations) to perform the CSL task. We apply our approach to three different datasets (two grounded language datasets and a robotic dataset) and observe that (1) One-Hot, GloVe, and pretrained BERT representations are less efficient when compared to representations obtained from fine-tuned BERT. (2) ESN online with final learning (FL) yields superior performance over ESN online continual learning (CL), offline learning, and LSTMs, indicating the more biological plausibility of ESNs and the cognitive process of sentence reading. (2) An LSTM with fewer hidden units showcases higher performance for small datasets, but an LSTM with more hidden units is needed to perform reasonably well on larger corpora. (4) ESNs demonstrate better generalization than LSTM models for increasingly large vocabularies. Overall, these models are able to learn from scratch to link complex relations between words and their corresponding meaning concepts, handling polysemous and synonymous words. Moreover, we argue that such models can extend to help current human-robot interaction studies on language grounding and better understand children's developmental language acquisition.

4.1.1.3 Highlights

This work compares extensively the performance of reservoir and LSTM networks (including random LSTMs), with various word embedding conditions. It shows how well reservoirs can generalize compared to LSTMs when learning on small datasets. Thus, interesting in the context of language acquisition modelling. It contrasts these architectures on multiple language & robotics datasets with different kinds of complexity (e.g. vocabulary size, length of sentences). It is a good ground for future neural network comparisons that will be made during the Research Program.

4.1.2 Hierarchical language processing: from speech to semantic labels

4.1.2.1 Context

This study [Pedrelli & Hinaut 2022] was done during the Post-Doc of Luca Pedrelli (2019-2020). Figures 5.3 and 5.4 in Chapter 5 summarize the main model and main qualitative results.

Luca Pedrelli and Xavier Hinaut (2022) Hierarchical-Task Reservoir for Online Semantic Analysis From Continuous Speech. IEEE Transactions on Neural Networks and Learning Systems, vol. 33, no. 6, pages 2654–2663.

- Journal published version:
  https://ieeexplore.ieee.org/document/9548713
- Open access / HAL version:
  https://hal.inria.fr/hal-03031413
- DOI: 10.1109/tnnls.2021.3095140
- Supplementary data:
  Directly available in the preprint.

4.1.2.2 Abstract

In this article, we propose a novel architecture called hierarchical-task reservoir (HTR) suitable for real-time applications for which different levels of abstraction are available. We apply it to semantic role labeling (SRL) based on continuous speech recognition. Taking inspiration from the brain, this demonstrates the hierarchies of representations from perceptive to integrative areas, and we consider a hierarchy of four subtasks with increasing levels of abstraction (phone, word, part-of-speech (POS), and semantic role tags). These tasks are progressively learned by the layers of the HTR architecture. Interestingly, quantitative and qualitative results show that

the hierarchical-task approach provides an advantage to improve the prediction. In particular, the qualitative results show that a shallow or a hierarchical reservoir, considered as baselines, does not produce estimations as good as the HTR model would. Moreover, we show that it is possible to further improve the accuracy of the model by designing skip connections and by considering word embedding (WE) in the internal representations. Overall, the HTR outperformed the other state-of-the-art reservoir-based approaches and it resulted in extremely efficient with respect to typical recurrent neural networks (RNNs) in deep learning (DL) [e.g., long short term memory (LSTMs)]. The HTR architecture is proposed as a step toward the modeling of online and hierarchical processes at work in the brain during language comprehension.

4.1.2.3 Highlights

We propose a new kind of deep reservoir architecture: the Hierarchical-Task Reservoir (HTR). This architecture proposes to solve a hierarchical task with each reservoir performing a sub-task. We created a challenging task by taking the well-known challenging TIMIT speech dataset: the layers of the reservoirs need to perform online phoneme recognition, word recognition, Part-of-Speech (POS) recognition and Semantic Role Labelling (SRL). We demonstrate that the architecture is able to solve the task with good performance. **It is important for the project** as the same architecture will be extended in WP1 and serve as a reference throughout the whole project. Interestingly, the architecture can represent all levels of abstraction in parallel during the processing of the utterance: it does not have to wait until the end of the sentence, it provides online predictions.

4.2 Songbird sensorimotor learning

4.2.1 Generating realistic sounds with low-dimensional space

4.2.1.1 Context

This study was done during the PhD thesis of Silvia Pagliarini (2017-2021) co-supervised with Arthur Leblois. It was performed with the help of Nathan Trouvain during his MSc. internship with me (spring 2020) and then as an engineer (2020-2022) again with me. An interesting qualitative result is presented in Figure 5.6 of Chapter 5.

Silvia Pagliarini, Nathan Trouvain, Arthur Leblois* and Xavier Hinaut* (2021) What does the Canary Say? Low-Dimensional GAN Applied to Birdsong. HAL preprint hal-03244723. *Equal contribution

- Open access / HAL version:
  https://hal.inria.fr/hal-03244723
- Supplementary data:
  Directly available in the preprint.

4.2.1.2 Abstract

The generation of speech, and more generally complex animal vocalizations, by artificial systems is a difficult problem. Generative Adversarial Networks (GANs) have shown very good abilities for generating images, and more recently sounds. While current GANs have high-dimensional latent spaces, complex vocalizations could in principle be generated through a low-dimensional latent space, easing the visualization and evaluation of latent representations. In this study, we aim to test the ability of a previously developed GAN, called WaveGAN, to reproduce canary syllables while drastically reducing the latent space dimension. We trained WaveGAN on a large dataset of canary syllables (16000 renditions of 16 different syllable types) and varied the latent space dimensions from 1 to 6. The sounds produced by the generator are evaluated using a RNN-based classifier. This quantitative evaluation is paired with a qualitative evaluation of the GAN productions across training epochs and latent dimensions. Altogether, our results show that a 3-dimensional latent space is enough to produce all syllable types in the repertoire with a quality often indistinguishable from real canary vocalizations. Importantly, we show that the 3-dimensional GAN generalizes by interpolating between the various syllable types. We rely on UMAP [McInnes *et al.* 2018] to qualitatively show the similarities between training and generated data, and between the generated syllables and the interpolations produced. We discuss how our study may provide tools to train simple models of vocal production and/or learning. Indeed, while the RNN-based classifier provides a biologically realistic representation of the auditory network processing vocalizations, the small dimensional GAN may be used for the production of complex vocal repertoires.

4.2.1.3 Highlights

We managed to train a GAN to generate raw sounds (WaveGAN) with a large dataset of canary syllables (16000 renditions) and to constrain the latent space to small dimensions (from 1 to 6). The sounds produced by the generator were identified and evaluated by a reservoir-based classifier trained on the same dataset. We also performed qualitative evaluation of the GAN outputs (using UMAP) across

GAN training epochs and latent dimensions. Uniform Manifold Approximation and Projection (UMAP) representations show the similarities between the training data and the generated data, and between the generated syllables and the interpolations produced. By exploring the latent representations of syllable types, we showed that they form well identifiable subspaces of the latent space. **It is important for the project**, because it shows that we are able to use reservoirs to evaluate noisy generations of GAN data and give qualitative evaluations with dimension reduction methods (UMAP). Moreover, it will serve as a complementary method to human vocal tract models for sound generations (when trained on speech).

4.2.2 Learning canary syllables with a simple Hebbian rule

4.2.2.1 Context

This study was done during the PhD thesis of Silvia Pagliarini (2017-2021) co-supervised with Arthur Leblois. It was performed with the help of Nathan Trouvain during his engineer position (2020-2022) with me. An interesting result is presented in Figure 5.7 of Chapter 5.

Silvia Pagliarini, Arthur Leblois* and Xavier Hinaut* (2021) Canary Vocal Sensorimotor Model with RNN Decoder and Low-dimensional GAN Generator. In ICDL 2021 - IEEE International Conference on Development and Learning, Beijing, China. *Equal contribution

- Conference published version:
  https://ieeexplore.ieee.org/abstract/document/9515607
- Open access / HAL version:
  https://hal.inria.fr/hal-03482372
- DOI: 10.1109/ICDL49984.2021.9515607

4.2.2.2 Abstract

Songbirds, like humans, learn to imitate sounds produced by adult conspecifics. Similarly, a complete vocal learning model should be able to produce, perceive and imitate realistic sounds. We propose (1) to use a low-dimensional generator model obtained from training WaveGAN on a canary vocalizations, (2) to use a RNN-classifier to model sensory processing. In this scenario, can a simple Hebbian learning rule drive the learning of the inverse model linking the perceptual space and the motor space? First, we study how the motor latent space topology affects the learning process. We then investigate the influence of the learning rate and of the motor latent space dimension. We observe that a simple Hebbian rule is able to drive the learning of realistic sounds produced via a low-dimensional GAN.

4.2.2.3 Highlights

This study is **important for the project** because it shows that we are able to have a sensorimotor model with a complete loop with the environment by generating real qualitative sounds and not just spectrograms. Moreover, this model is able to learn nearly all syllables with a simple Hebbian learning rule, which is very promising for the Research Program when we will be using more advanced rules based on reinforcement learning.

4.3 Prefrontal cortex & working memory

4.3.1 Building line-attractors: training reservoirs to Gate

4.3.1.1 Context

This study was started during Anthony Strock's master intership (spring 2017) and realized during his PhD thesis (2017-2020) co-supervised with Nicolas Rougier. Figure 5.8 in Chapter 5 is summarizing the main model.

Anthony Strock, Xavier Hinaut* and Nicolas P. Rougier* (2020) A Robust Model of Gated Working Memory. Neural Computation, vol. 32, no. 1, pages 153–181. *Equal contribution

- Journal published version:
  https://direct.mit.edu/neco/article-abstract/32/1/153/95568/A-Robust-Model-of-Gated-Working-Memory
- Open access / HAL version:
  https://hal.inria.fr/hal-02371659/
- DOI: 10.1162/neco_a_01249
- Supplementary data:
  https://hal.archives-ouvertes.fr/hal-02371659/file/NECO-05-19-3483-supplementary.pdf

4.3.1.2 Abstract

Gated working memory is defined as the capacity of holding arbitrary information at any time in order to be used at a later time. Based on electrophysiological recordings, several computational models have tackled the problem using dedicated and explicit mechanisms. We propose instead to consider an implicit mechanism based on a random recurrent neural network. We introduce a robust yet simple reservoir model of gated working memory with instantaneous updates. The model is able to store an arbitrary real value at random time over an extended period of time. The dynamics

of the model is a line attractor that learns to exploit reentry and a nonlinearity during the training phase using only a few representative values. A deeper study of the model shows that there is actually a large range of hyperparameters for which the results hold (e.g., number of neurons, sparsity, global weight scaling) such that any large enough population, mixing excitatory and inhibitory neurons, can quickly learn to realize such gated working memory. In a nutshell, with a minimal set of hypotheses, we show that we can have a robust model of working memory. This suggests this property could be an implicit property of any random population, that can be acquired through learning. Furthermore, considering working memory to be a physically open but functionally closed system, we give account on some counterintuitive electrophysiological recordings.

4.3.1.3 Highlights

We were able to model working memory mechanisms by training a simple random recurrent network to gate information for a long period of time. There are four main findings: (1) the model extends the memory capacity of reservoirs with only one continuous output node, (2) the model is very robust against perturbations while not being hand-crafted, (3) we can fit different kinds of dynamics observed in experimental data by simply changing a meta-parameter, (4) we explain the model by deriving a minimal model based on 3 neurons that produce equivalent behavior even in case of lesion. Such "gated reservoir" can be used as a new tool by the reservoir community. **This is important for the project** because this new tool will be used in the project (e.g. to stabilize representations for long-time periods). Moreover, it shows that our expertise in reservoir computing enabled us to extend the memory capacity of reservoir for long-time dependencies. Such expertise will be useful to create new reservoir computing tools during this project.

4.4 Bringing together tools for reservoir exploration

During my PhD, I had the chance to be part of the european FP7 *Organic* project with the founders of Reservoir Computing. *Oger*[1] library was developed during the project but discontinued shortly afterwards. It was nice to have such a library and being able to discuss and share code easily with my PhD partner Pierre Enel. Since then, I was looking for a similar one, but I did not find one that was fitted my needs. That's why I started developing ReservoirPy, which was then greatly extended and reshaped by Nathan Trouvain when he was engineer the team. We have now a well documented and flexible tool to prototype quickly reservoir architectures. We also developed new tools to analyse and visualize the internal dynamics of reservoirs to enable comparisons with other kinds of networks such as LSTMs.

[1] I host on my GitHub the last version of Oger which was available on BitBucket: https://github.com/neuronalX/Oger

4.4.1 ReservoirPy: reservoirs in few lines of code

4.4.1.1 Context

This work [Trouvain & Hinaut 2022] was done during the engineer position of Nathan Trouvain (2020-2022). It follows the development that Nathan started when he was in MSc. internship again with me [Trouvain *et al.* 2020] (spring 2020). It was performed with the help of Nicolas Rougier. In Figure 5.10 of Chapter 5 an example of hyperparameter search done with ReservoirPy and *hyperopt* library is proposed.

Nathan Trouvain and Xavier Hinaut (2022) reservoirpy: A Simple and Flexible Reservoir Computing Tool in Python. hal-03699931 preprint.

- Open access / HAL version:
  https://hal.inria.fr/hal-03699931

4.4.1.2 Abstract

This paper presents *reservoirpy*, a Python library for Reservoir Computing (RC) models design and training, with a particular focus on Echo State Networks (ESNs). The library contains basic building blocks for a large variety of recurrent neural networks defined within the field of RC, along with both offline and online learning rules. Advanced features of the library enable compositions of RC building blocks to create complex "deep" models, delayed connections between these blocks to convey feedback signals, and empower users to create their own recurrent operators or neuronal connections topology. This tool is solely based on Python standard scientific packages such as numpy and scipy. It improves RC time efficiency with parallelism using joblib package, making it accessible to a large academic or industrial audience even with a low computational budget. Source code, tutorials and examples from the RC literature can be found at https://github.com/reservoirpy/reservoirpy while documentation can be found at https://reservoirpy.readthedocs.io/en/latest/?badge=latest

4.4.1.3 Highlights

The ReservoirPy library will be **used during the Research Program** to quickly prototype reservoir models, but also to share the models and gather a community around the project. It will also enable to compare various features and learning rules.

4.4.2 Easily building complex architectures

4.4.2.1 Context

This work [Trouvain *et al.* 2022] was done during the engineer position of Nathan Trouvain (2020-2022). It was performed with the help of Nicolas Rougier.

Nathan Trouvain, Nicolas P. Rougier and Xavier Hinaut (2022) Create Efficient and Complex Reservoir Computing Architectures with ReservoirPy. In SAB 2022 - FROM ANIMALS TO ANIMATS 16: The 16th International Conference on the Simulation of Adaptive Behavior, Cergy-Pontoise / Hybrid, France.

- Journal published version:
  https://link.springer.com/chapter/10.1007/978-3-031-16770-6_8
- Open access / HAL version:
  https://hal.inria.fr/hal-03761440
- DOI: 10.1007/978-3-031-16770-6_8
- Supplementary data:
  Directly available in the preprint.

4.4.2.2 Abstract

Reservoir Computing (RC) is a type of recurrent neural network (RNNs) where learning is restricted to the output weights. RCs are often considered as temporal Support Vector Machines (SVMs) for the way they project inputs onto dynamic non-linear high-dimensional representations. This paradigm, mainly represented by Echo State Networks (ESNs), has been successfully applied on a wide variety of tasks, from time series forecasting to sequence generation. They offer de facto a fast, simple yet efficient way to train RNNs. We present in this paper a library that facilitates the creation of RC architectures, from simplest to most complex, based on the Python scientific stack (NumPy, Scipy). This library offers memory and time efficient implementations for both online and offline training paradigms, such as FORCE learning or parallel ridge regression. The flexibility of the API allows to quickly design ESNs including re-usable and customizable components. It enables to build models such as DeepESNs as well as other advanced architectures with complex connectivity between multiple reservoirs with feedback loops. Extensive documentation and tutorials both for newcomers and experts are provided through GitHub and ReadTheDocs websites. The paper introduces the main concepts supporting the library, illustrated with code examples covering popular RC techniques from the literature. We argue that such flexible dedicated library will ease the creation of more advanced architectures while guarantying their correct implementation and reproducibility across the RC community.

4.4.2.3 Highlights

The inclusion of tools to design complex architectures available within the new version of the ReservoirPy library will enable us to **explore a variety of topology of reservoir networks during the Research Program**.

4.4.3 Diving into reservoirs and LSTMs generalization

4.4.3.1 Context

This study [Variengien & Hinaut 2020] was performed with Alexandre Variengien during his BSc internship (spring 2020).

Alexandre Variengien and Xavier Hinaut (2020) A journey in ESN and LSTM visualisations on a language task. arXiv:2012.01748.

- Open access / HAL version:
  https://hal.inria.fr/hal-03030248
- Open access / arXiv version:
  https://arxiv.org/abs/2012.01748
- Supplementary data:
  Directly available in the preprint.

4.4.3.2 Abstract

Echo States Networks (ESN) and Long-Short Term Memory networks (LSTM) are two popular architectures of Recurrent Neural Networks (RNN) to solve machine learning task involving sequential data. However, little have been done to compare their performances and their internal mechanisms on a common task. In this work, we trained ESNs and LSTMs on a Cross-Situationnal Learning (CSL) task. This task aims at modelling how infants learn language: they create associations between words and visual stimuli in order to extract meaning from words and sentences. The results are of three kinds: performance comparison, internal dynamics analyses and visualization of latent space. (1) We found that both models were able to successfully learn the task: the LSTM reached the lowest error for the basic corpus, but the ESN was quicker to train. Furthermore, the ESN was able to outperform LSTMs on datasets more challenging without any further tuning needed. (2) We also conducted an analysis of the internal units activations of LSTMs and ESNs. Despite the deep differences between both models (trained or fixed internal weights), we were able to uncover similar inner mechanisms: both put emphasis on the units encoding aspects of the sentence structure. (3) Moreover, we present Recurrent States Space Visualisations (RSSviz), a method to visualize the structure of latent state space of RNNs, based on dimension reduction (using UMAP). This technique

enables us to observe a fractal embedding of sequences in the LSTM. RSSviz is also useful for the analysis of ESNs (i) to spot difficult examples and (ii) to generate animated plots showing the evolution of activations across learning stages. Finally, we explore qualitatively how the RSSviz could provide an intuitive visualisation to understand the influence of hyperparameters on the reservoir dynamics prior to ESN training.

4.4.3.3 Highlights

This study extended results of [Juven & Hinaut 2020] and provided earlier results that have been extended in [Oota *et al.* 2022] (see Subsection 4.1.1). It provides a qualitative analysis to seek why reservoirs generalize with less data than LSTMs. It studies the internal dynamics of both networks in various ways: it compares their "richness" *vs.* specificity, provides UMAP [McInnes *et al.* 2018] representations of internal trajectories and UMAP representations of the changes occurring during training. It also gives an analysis of the effect of hyperparameters on unit activations and statespace representations. This work provides **analysis tools that will be useful during the Research Program**.

Chapter 5

Research Program

Contents

> I stood still, my whole attention fixed upon the motions of her fingers. Suddenly I felt a misty consciousness as of something forgotten — a thrill of returning thought; and somehow the mystery of language was revealed to me. I knew then that w-a-t-e-r meant the wonderful cool something that was flowing over my hand. The living word awakened my soul, gave it light, hope, set it free!
>
> *The Story of My Life*
> *Helen A. Keller.*

5.1 Hierarchical reservoirs to model language processing and production

Abstract. Language involves several abstraction levels of hierarchy. Most models focus on a particular level of abstraction making them unable to model bottom-up and top-down processes. Moreover, we do not know how the brain grounds symbols to perceptions and how these symbols emerge throughout development. Experimental evidence suggests that perception and action shape one-another (e.g. motor areas activated during speech perception) but the precise mechanisms involved in this action-perception shaping at various levels of abstraction are still largely unknown. I propose to create a new generation of neural-based computational models of language processing and production: i.e. to (1) use biologically plausible learning mechanisms; (2) create novel sensorimotor mechanisms to account for action-perception shaping; (3) build hierarchical models from sensorimotor to sentence level; (4) embody such models in robots in order to ground semantics.

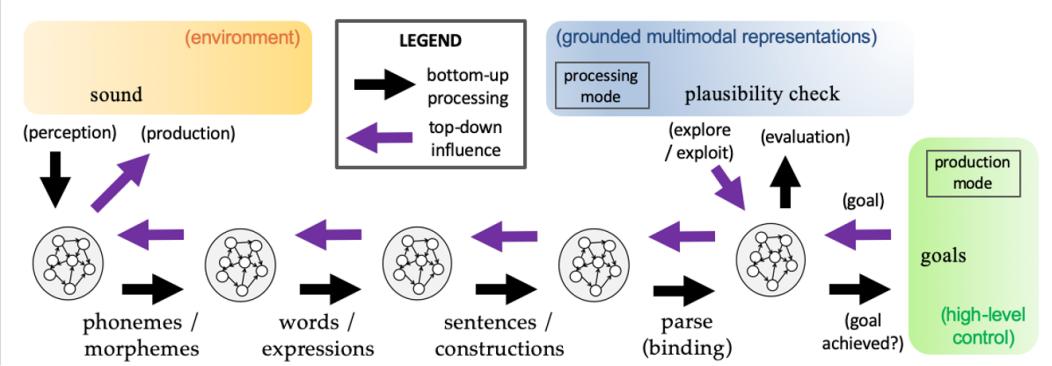

Figure 5.1: General target architecture of the project.

5.1.1 Scientific context and motivation

After brain strokes (e.g. causing aphasia), it is not clear how the brain manages to reorganize language functions. Computational neural models would be crucial to provide a deeper understanding of such language functions. Recently Deep Learning (DL) networks have created a breakthrough in image and speech recognition, and Natural Language Processing (NLP) methods. However, no equivalent breakthrough happened towards the understanding of how the brain performs similar functions. This breakthrough did not happen because Deep Learning does not yet reproduce learning mechanisms nor the dynamics of the brain. Thus, we still **lack the key neuronal mechanisms needed to properly model the (hierarchies of) functions in language perception and production**.

The brain needs to parse incoming stimuli and learn from them incrementally, it cannot unfold time like classical DL algorithms such as Back-propagation through

time (BPTT). This would be equivalent to have a virtual copy of our brain for each time step and use the last hundreds virtual brain copies to learn long-time dependencies. To model language processes of healthy or lesioned brains some models try to reproduce the behaviour of brain dynamics: e.g. with Event-Related-Potential (ERP) data using backpropagation [Brouwer & Hoeks 2013, Brouwer *et al.* 2017] or recurrence analysis [beim Graben & Hutt 2015]. However, such models lack explanatory power demonstrating the causes of such observed dynamics: i.e. what is computed and why is it computed – for which purpose? Other models have core mechanisms that are well engineered in order to perform the task and do not reflect biological mechanisms. **We need more biologically plausible learning mechanisms favoring emergence while producing causal explanations of the experimental data modelled.**

There is converging evidence that language production and comprehension are not separated processes in a "modular mind", they are rather interwoven, and this interweaving is what enables people to predict themselves and each other [Pickering & Garrod 2013]. Interweaving of action and perception is important because it allows a learning agent (or a baby) to learn from its own actions: for instance, by learning the perceptual consequences (e.g. the heard sounds) of its own actions (e.g. vocal productions) during babbling. Thus, the agent will learn in a **self-supervised** way instead of relying only on supervised learning, which in contrast, imply non-biological teacher signals cleverly designed by the modeller. Self-supervised learning is fundamental for developmental processes such as babbling. Schwartz et al. [Schwartz *et al.* 2012] propose that perception and action are co-structured in the course of speech development: gestures are perceptually-shaped, they form a perceptuo-motor unit. **A clear neuronal model explaining which are the mechanisms shaping these perceptuo-motor units through development is missing.**

In order to obtain new emergent representations of morphemes, words and sentences we cannot rely on engineered ones (e.g. word embeddings such as Word2Vec [Mikolov *et al.* 2013] or BERT [Devlin *et al.* 2018]). We need to obtain **emergent** action-perception representations through perceptuo-motor mechanisms. The existence of sensorimotor (i.e. mirror) neurons at abstract representation levels (called action-perception circuits [Pulvermüller & Fadiga 2010]), jointly with the perceptuo-motor shaping of sensorimotor gestures, suggest the existence of **similar action-perception mechanisms implemented at different levels of hierarchy.**

Christiansen & Chater propose that the brain is in the *Now or Never Bottleneck* problem [Christiansen & Chater 2016] when processing a stimulus (e.g. an utterance): it is forced to extract the necessary information as soon as possible, otherwise the information will be lost. Thus, the **rich perceptual input needs to be recoded as it arrives in order to capture the key elements of the sensory information** [Christiansen *et al.* 2016]. These compressed (or "chunked") representations are abstractions of inputs (filtering out the details) rather than predictions encoding all the fluctuations of fast incoming inputs. Memory limitations also apply

to these recoded representations; hence the brain needs to chunk the compressed representations into *"multiple levels of representation of increasing abstraction in perception, and decreasing levels of abstraction in action"* [Christiansen *et al.* 2016]. Therefore, each sequence of chunks at one level will be encoded as a single chunk to a higher level. In summary, they suggest the brain must implement a hierarchical "Chunk and Pass" mechanism to solve the "Now or Never Bottleneck" problem.

Importantly, a language model needs a way to acquire the semantics of the (symbolic) perceptuo-motor gestures and of the more abstract representations, otherwise it would consider only morphosyntactic and prosodic features of language. These symbolic gestures, i.e. signs, need to be grounded to the mental concept, i.e. signified, they are representing. Several theories and robotic experiments give examples of how symbols could be grounded or how symbols could emerge [Taniguchi *et al.* 2016]. These are important conceptual questions for AI (Artificial Intelligence) in robotics. It is also crucial to understand how the brain solve these problems. However, current neurocomputational models aiming to explain brain processes [Garagnani *et al.* 2008, Garagnani & Pulvermüller 2016, Brouwer *et al.* 2017] are not grounded[1]. We need mechanisms that start from raw sensory perception and raw motor commands in order to let emerge plausible representations through development, instead of arbitrary representations. We target to embody models into **robots that will developmentally ground language from morphemes to sentences**. The grounding of semantics should come from the robot experiencing the world through its interactions with humans and the physical world.

5.1.2 Objectives and research hypothesis

The aim of the project is to build a dynamic neuronal model of language processing and production: the model should be dynamic, grounded, hierarchical and use action-perception mechanisms.

One of the long-term ambitions of the project is to make a biologically plausible model of sentence processing and production. Such model could be used to fit experimental data of healthy and pathological language functions. Moreover, the project will explore how the model could be embodied [Pulvermüller 2013] in robots in order to model (among other things) how the brain of children learns to ground, in a developmental scheme, the semantics of various levels of symbol abstraction. The model will have to deal with continuous stimuli (speech) and produce continuous actions (vocalizations). Thus, on one side, it will deal with unsegmented streams from (bottom-up) perceptual categories (e.g. phonemes), sequences of abstracted categories up to sentences, and conversely on another side, it will produce sentences going from abstract representations to the sequence of syllables. An important

[1] Some authors seem to use "neuroanatomically grounded" for neurocomputational models that model neuroanatomy [Garagnani *et al.* 2008]. I do not use the term "grounded" in this sense.

novelty is that the goal is to model this bottom-up and top-down processes with action-perception mechanisms.

The developed models will be dynamic: time will not be chunked (i.e. segmented) or unfolded during simulations, trained online (training will happen online within the simulation processes) and anytime (if the simulation is stopped at any time it will give a partial result, e.g. partial thematic roles available during the processing of sentences). Some of these features are already part our previous studies, so the challenge will be to create some features (e.g. chunking) while keeping the previous ones.

The final model will be composed of several sub-models which will tackle four challenges: (1) Dynamic and developmental models: Which is the combination of learning rules that enable generic recurrent neural networks (RNN), such as reservoirs, to learn incrementally, from temporally distant rewards and build representations upon one another in a developmental way? (2) Action-Perception generic mechanism: How to make a generic action-perception mechanism that (i) would enable action and perception to shape one another, (ii) while allowing to bootstrap the development of representations from raw sound percepts, and (iii) which could be stacked as layers of a hierarchy? (3) Hierarchical: from sensori-motor learning to sentence comprehension: How to create (i) a layered architecture with reservoirs that is working functionally as a hierarchy (i.e. each layer as a specific function and abstraction level), (ii) such that it allows bottom-up and top-down processes to flow, while enabling multistable representations [Kelso *et al.* 1995, Sterzer *et al.* 2009]. (4) Grounded in virtual agents and physical robots: How to integrate semantics from other modalities into the active-perceptive hierarchical model obtained so far?

Obtaining a functional hierarchy is not only useful to abstract symbols from raw perceptions (i.e. let them emerge), but also to create action-perception couplings with the environment at different levels of abstraction, in order to be able to reproduce (i.e. to imitate with motor commands) past perceived stimuli.

A new kind of bio-plausible mechanism is needed to achieve this challenge: we need to think outside of the common input-output black box training of neural networks and see how they can be trained jointly including the environment. Sensorimotor models and Generative Adversarial Networks (GANs) are interesting bases to start, but we need to go further. Thus, part of the objectives, is to discover alternative solutions to the input-output mapping black box dominant paradigm. Even if some methods, such as GANs, are not "one-way only" and enable the system to be partly self-supervised, by design each "box" (i.e. module) has its own input and output. After training, each "box" can be used independently from other modules. With this project, we want to enable different parts of the global system to truly interact dynamically, like a dynamical system: by having a motor/generative module (that produces phonemes/words/...) able to bias the perceptual module. Importantly, part of the general aim is to enable such skewing mechanism to happen hierarchically: more abstract layers could bias lower layers.

One of the long-term goals of the project is to build a model that could learn to understand utterances by exploring which meanings the morphemes, words, expres-

sions, etc. could have within the context of a sentence. This assumes that sentence comprehension is not a passive process simply chunking information at different levels of abstraction, but on the contrary, it is an active process where the hearer tries to infer the meaning of the sentence. A sentence may not be understandable based only on the most probable on-going parse: garden path sentences (e.g. "The horse raced past the barn fell.") exemplify such need for active exploration. Testing this ability to chunk multi-word expression will be especially interesting and we could link model activities to functional Magnetic Resonance Imaging (fMRI) studies looking at such expressions [Bhattasali *et al.* 2019].

Additionally, it is important to keep models as generic as possible, in order to prevent from adding unnecessary linguistic a priori knowledge. Generic models will have a greater impact, opening potential adaptations to non-linguistic tasks. In particular, we do not want to predefine connections between symbolic components: we want the symbols (or "perceptuo-motor gestures") to be dynamically connected in a generic structure, and not to appear as a result of an engineered connectivity or mechanisms.

Overall, this project aims to initiate a paradigm shift in the design of linguistic cognition and how it develops and demonstrate the efficiency of hierarchical active perception mechanisms in noisy conditions of real-world applications (e.g. Human-Robot Interaction).

5.1.2.1 Expected results

Final demonstration The final language model will be implemented in an interactive scenario between a human user and a robot. Users could interact with the robot through speech commands, and the robot will perform the actions. Users could talk different languages to the robot. During the interaction users could see a schematized version of the model activities going-on with a screen. More details concerning a particular area could be obtained: e.g. see potential ambiguous words not fully recognized by the robot. Finally, users could see on the screen the replay of the developmental stages the model has passed by.

5.1.3 Position of the project as it relates to the state of the art

The way we learn to ground utterances to meaningful representations is a complex process that involves to link many sub-processes of different nature. Barsalou [Barsalou 2008] proposes that "language provides an excellent domain in which to combine symbolic operations [Sun & Alexandre 2013], statistical processing and grounding".

Several mechanisms are discussion topics regarding biological plausibility, however relying on back-propagation is often not considered plausible, especially if the gradient needs to go backwards through several layers. Back-propagation through time (BPTT) makes the implausibility a step further, as it needs to unfold time, which means to virtualize it as a spatial dimension in order to train a recurrent neu-

ral network (RNN). In fact, since a decade my studies are mainly focused on RNNs that do not involve the unfolding of time (e.g. BPTT) such as Echo State Networks (ESNs) [Jaeger & Haas 2004]. Most parts of the weights of such models are not trained, this make a difference on biological plausibility compared to other classical algorithms: one could consider that at a short timescale the weights do not change in the reservoir [Lukoševičius & Jaeger 2009], or that optimisation of hyperparameters of the reservoir actually corresponds to what would be obtained with homeostasis rules such as intrinsic plasticity mechanisms [Steil 2007, Schrauwen *et al.*]. On the contrary, an important number of previously developed neural networks models on sensorimotor learning use feed-forward neural networks trained with backpropagation: this is a common way to escape the problem of representing time as such.

Recently, deep learning networks [Graves *et al.* 2013, Cho *et al.* 2014b, Chung *et al.* 2016, Devlin *et al.* 2018, Luong *et al.* 2015] have created a breakthrough in object and speech recognition. Latent representations of word and sentences, such as Word2Vec [Mikolov *et al.* 2013], and subsequent developments, such as transformers BERT, RoBERTa, GPT-2 etc. [Vaswani *et al.* 2017, Devlin *et al.* 2018, Liu *et al.* 2019, Radford *et al.* 2019, Raffel *et al.* 2020], enabled important progress on language modelling and natural language processing (NLP).

However, no equivalent breakthrough happened towards the understanding of how the brain performs similar functions. Recent studies show that neural networks are able to predict human brain responses elicited during reading tasks, e.g. magneto-encephalography (MEG) recordings [Caucheteux & King 2021]. However, when the algorithms are trained on language modeling, only the middle layers become increasingly similar to the late responses of the language network in the brain. Additionally, these models end up with a lot of trained parameters which make them too complex to analyze, and able to extract equivalent cognitive mechanisms performed by the brain.

Predictive coding and active inference are important mainstream approaches in several domains [Friston 2018][Pitti *et al.* 2020]. There are of course interesting common features with our project, as for other general theories. However, we want to propose mechanisms which do not rely mainly on *predictions* or *prediction error* as key mechanisms. As we have shown previously, predictions can appear as a by-product of learning [Hinaut & Dominey 2013]. We propose a challenging vision which is rather bottom-up than top-down: find mechanisms that are more "data-oriented" and fundamentally anchored in what is known form sensory perception and production rather than more abstract theories.

Sentence processing and production models. Early neuronal (i.e. connectionist) language processing models used backpropagation to predict the next word in a sentence [Elman 1990] or simple case-role assignment. Many subsequent models, e.g. of language production [Chang 2002], continued to use backpropagation as the main learning mechanism. Such models are interesting to fit with developmental data on language acquisition or time-reading data, but they cannot model brain dynamics [Crocker *et al.* 2006]. More recently, some mod-

els targeted the modelling of ERP (Event-Related Potentials) components obtained from EEG [Brouwer & Hoeks 2013, Brouwer *et al.* 2017]. Interestingly, some models aimed at simulating brain activations with an action-perception point of view [Pulvermüller & Fadiga 2010], but the mechanisms proposed do not provide sufficient explanations on what the brain computes when processing or producing a sentence.

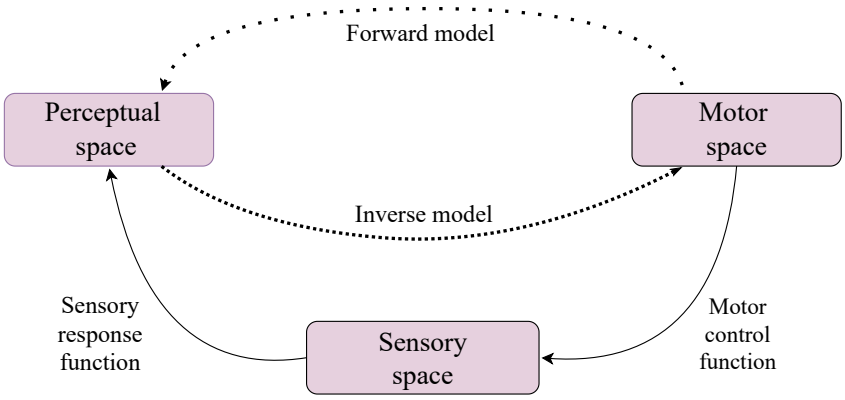

Figure 5.2: General archtitecture of a sensorimotor model. From our review [Pagliarini *et al.* 2021b].

Sensorimotor models. Recently, we made a review on sensorimotor models of vocal imitation [Pagliarini *et al.* 2021b]. Although these models were targeting similar functions, i.e. how humans or songbirds learn to vocalize, the resulting models were difficult to compare, because there were using too much different methods. Grouping them in three main spaces (motor, sensory and perceptual spaces) and associated functions allowed us to compare them. On Figure 5.2, one can see the "minimal" architecture that enables to focus on the core model parts. We describe briefly these core parts: the perceptual space represents the result of the sound transformation by the sensory response function; the motor space represents the motor commands before being transformed into sound by the motor control function [Tourville & Guenther 2011] (e.g. a human articulatory speech synthesizer model [Birkholz *et al.* 2006, Kröger & Bekolay 2019]; the sensory space is the sound in vocal imitation models; the inverse model allows to provide an appropriate motor command for a given perceptual goal (mapping perceptions to motor commands); the forward model (mapping motor commands to perceptions) describes a causal relationship between motor commands and their corresponding perceptual representations, i.e. it tries to predict the perceptions when a motor command is produced.

Perceptuo-motor shaping. The role of premotor cortex activation during speech perception has been discussed since a while [Meister *et al.* 2007] and is believed to enable for better representations of speech sounds, especially in noisy environment. The Perception-for-Action-Control Theory (PACT) [Schwartz *et al.* 2012]

highlights how speech percepts are related not only to sounds, but also to motor gestures: speech perception could be biased by articulatory invariant commands. Thus, syllables are perceptuo-motor in essence: perception-shapes-action (e.g. some abstract representation of motor gestures can be recovered to disambiguate perception) and action-shapes-perception (e.g. motor gestures are "selected for their functional and perceptual value for communication" [Schwartz *et al.* 2012]). An example is the fact that acoustic features can change suddenly when changing the jaw height or jaw cycle, thus producing phase transitions in the perceptuo-motor phase space diagram [Schwartz *et al.* 2012]. Bayesian computational models of PACT theory have been proposed [Moulin-Frier *et al.* 2015], but they do not enable to model brain processing at the mechanistic level. Moreover, such models do not consider the word nor sentence levels.

Grounding. One way to understand how symbols are grounded (or conversely how symbols emerge) is to use robots as models to see how to embody brain models of perception and action. In other words, we want to make these models interact with the world to ground/build themselves the symbols (as systems/agents with sensors and actuators which can perceive and act on the real world). Using robots to study language grounding, acquisition and development is a challenging research topic with several sub-fields. Cangelosi et al.[Cangelosi *et al.* 2010] have proposed an ambitious road-map plan for the integration of action and language through developmental robotics. Several studies attempted to tackle different sub-problems: word and syntax grounding with visual cues [Roy 2002], abstract word grounding [Stramandinoli *et al.* 2012], cross-situational learning [Taniguchi *et al.* 2017], grounded language acquisition [Dominey & Boucher 2005], symbol emergence [Taniguchi *et al.* 2016], semantic compositionality [Sugita & Tani 2005], origin of syntax with language games [Steels 1998].

Tani [Yamashita & Tani 2008] and others have been using robots to ground high-level cognition. Tani's work on hierarchical organization of motor actions is among the few models building a system that links low and high level temporal sensori-motor representations in a direct fashion. However, such approaches use learning methods that could be computationally costly and few provide developmental schemes which prevent them to scale, and it seems that none provide biologically plausible learning mechanisms. Moreover, one main limitation of such experiments on grounding with neural networks is that they use simple linguistic associations with hand-crafted preprocessing (e.g. usually grounding one perception to one word) because they do not have better options: they are lacking a hierarchical model representing the different levels of abstraction while processing sentences incrementally.

Modelling brain processes from raw acoustic signal up to language understanding is an ambitious long-term research project. There is no such multi-level and hierarchical model on language today, which crucially lacks for the neuro-linguistic and psycho-linguistic communities. If one considers a model at a single level (e.g. sentence level), one needs to make nearly arbitrary assumptions on how activi-

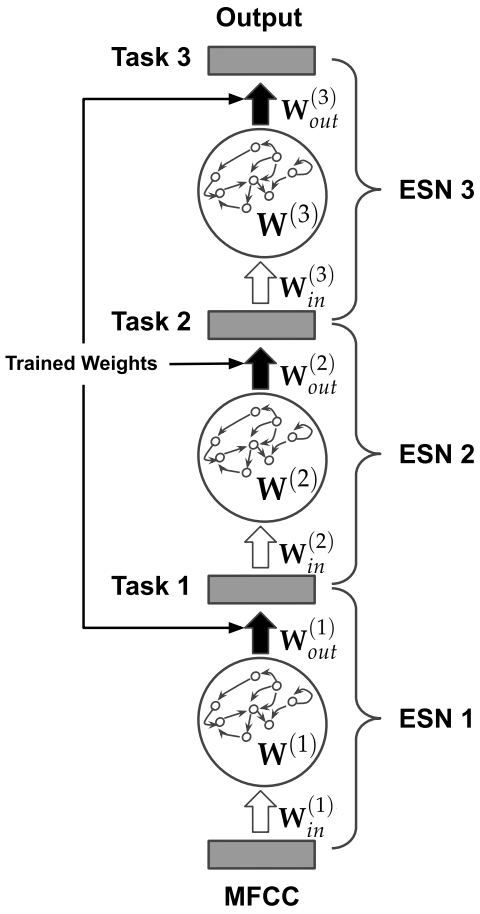

Figure 5.3: Hierarchical-Task Reservoir (HTR) architecture. Each ESN learns a different task. In [Pedrelli & Hinaut 2020] these tasks correspond to {PH, WD or POS}: obtaining the following flux of information MFCC→SP→PH→WD →POS. The architecture receives the features extracted from the speech signal with MFCC representations each 10 ms. The first layer is optimized for SP→PH (**Task 1**), the second layer is optimized for PH→WD (**Task 2**) and, finally, the third layer is optimized for WD→POS (**Task 3**). In [Pedrelli & Hinaut 2022] these tasks correspond to {PH, WD, POS or SRL}: we tested different kinds of architectures with different kind of representation (word embedding), including ones with skip connections from WD to SRL. Image from [Pedrelli & Hinaut 2020].

ties/information are represented (e.g. words are encoded as "grand-mother" neural assemblies). Such assumptions could be true, but there are much more chances that they are not, making the validity of the model questionable and limited to a narrow domain. Kröger et al. (2008) did a neurocomputational model of speech perception and production that spans on multiple levels, but only until phonemic map. Consequently, the hierarchical models will be designed in order to be adaptable to other hierarchical tasks like recognition and production of complex actions: my early work was on modelling abstract actions sequences.

5.1.3.1 Preliminary results

Our published works and ongoing works will serve as a basis for this project. In a first batch of studies [Pedrelli & Hinaut 2020, Pedrelli & Hinaut 2022], we propose a novel architecture called Hierarchical-Task Reservoir (HTR) suitable for real-time sentence parsing from continuous speech. Accordingly, we introduce a novel task that consists in performing anytime Semantic Role Labelling (SRL) from continuous speech. This HTR architecture is designed to address four classification sub-tasks (phones words, Part-of-Speech (POS) tags, SRL) with increasing levels of abstraction [Pedrelli & Hinaut 2022]. These tasks are performed by the consecutive layers of the HTR architecture. Interestingly, the results show that learning sub-tasks enforces better qualitative outputs [Pedrelli & Hinaut 2020, Pedrelli & Hinaut 2022] compared to a hierarchical reservoir predicting the same task at each layer [Triefenbach *et al.* 2013]. We compared HTR with a baseline hierarchical reservoir architecture and usual ESNs or LSTMs (Long-Short Term Memory networks) [Hochreiter & Schmidhuber 1997]. Moreover, we also performed a thorough experimental comparison with several architectural variants. Finally, the HTR with word embeddings and one skip connection (words->SRL) obtained the best performance.

In a second batch of studies [Dinh & Hinaut 2020, Juven & Hinaut 2020, Variengien & Hinaut 2020], we tackle the question of "Understanding the mechanisms enabling children to learn rapidly word-to-meaning mapping through cross-situational learning (CSL) in uncertain conditions". In particular, many models of language acquisition often look at the word level, and not at the full sentence comprehension level. We adapted our previous reservoir model [Hinaut & Dominey 2013] to learn to represent concepts instead of predicates (center of Figure 5.5). Using the co-occurrences between words and visual perceptions, the model learns to ground a complex sentence, describing a scene involving different objects, into a perceptual representation space (bottom of Figure 5.5). The reservoir processes sentences describing scenes and is trained to output the concepts given by the simulated vision module with online FORCE learning [Sussillo & Abbott 2009]. Evaluations of the model show its capacity to extract the semantics even if the concepts given as output often do not exactly correspond to the given sentence in input (i.e. CSL). Remarkably the model generalizes, on sentences describing one or two objects, only after a few hundred of partially described scenes. Furthermore, it handles polyse-

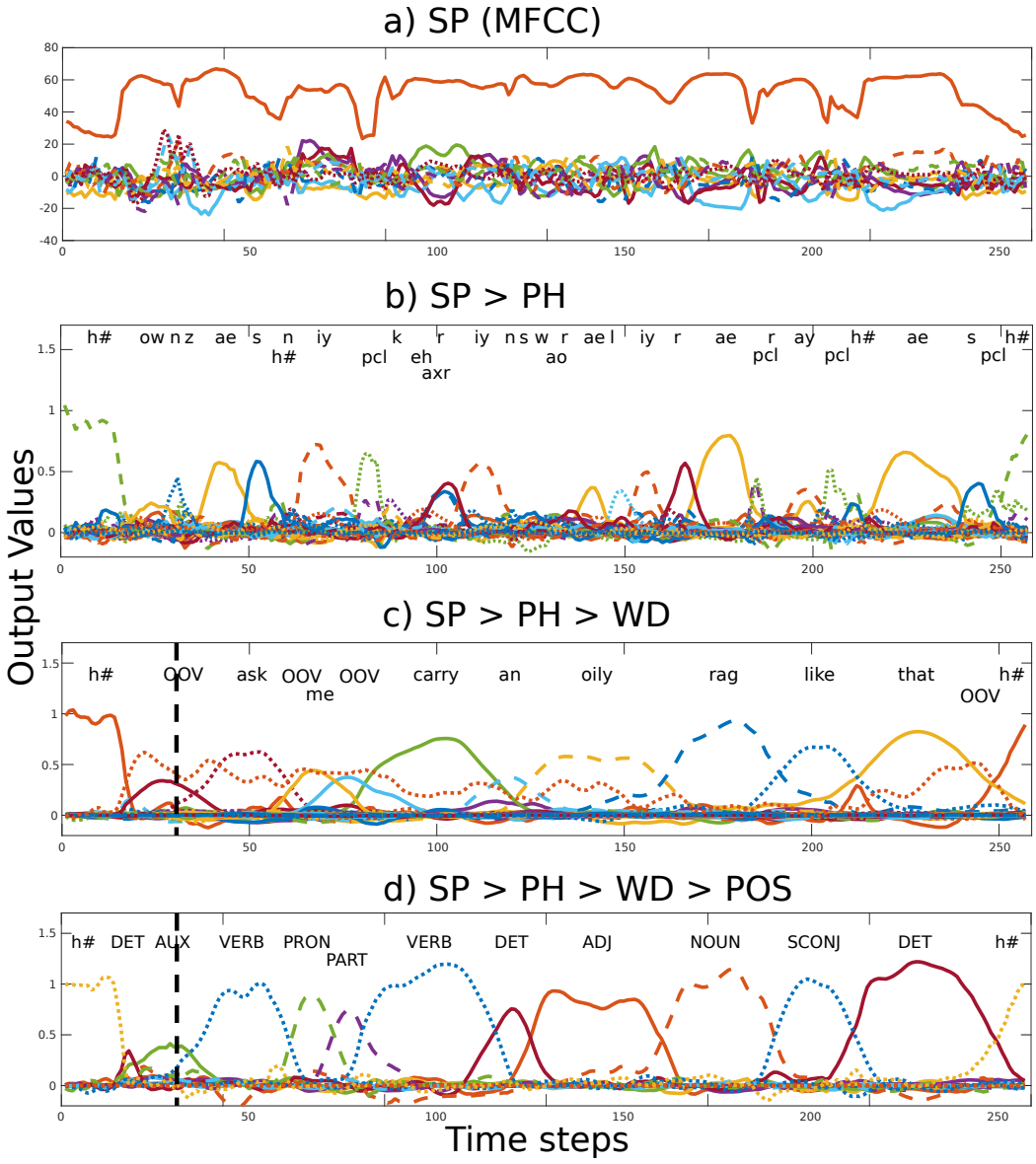

Figure 5.4: Input and outputs at different stages of a Hierarchital-Task Reservoir (HTR) performing 3 tasks – phone (PH), word (WD) and part-of-speech (POS). It is composed of 3 ESNs with the following flow: MFCC→SP→PH→WD →POS. The plot (a) shows the components of the MFCC computed from the input audio relative to the sentence *"Don't ask me to carry an oily rag like that"*. Plots (b), (c) and §d) show the output values of the the layers 1, 2 and 3 of the HTR architecture. The vertical dash line on bottom left indicates a "correction" that is made by the last ESN: input word is an OOV word, i.e. a word that is out of the top50 words in the corpus. The x-axis represents the time and the y-axis represents the values. At several time points, the label corresponding to the output with the maximum activation is indicated. Image from [Pedrelli & Hinaut 2020].

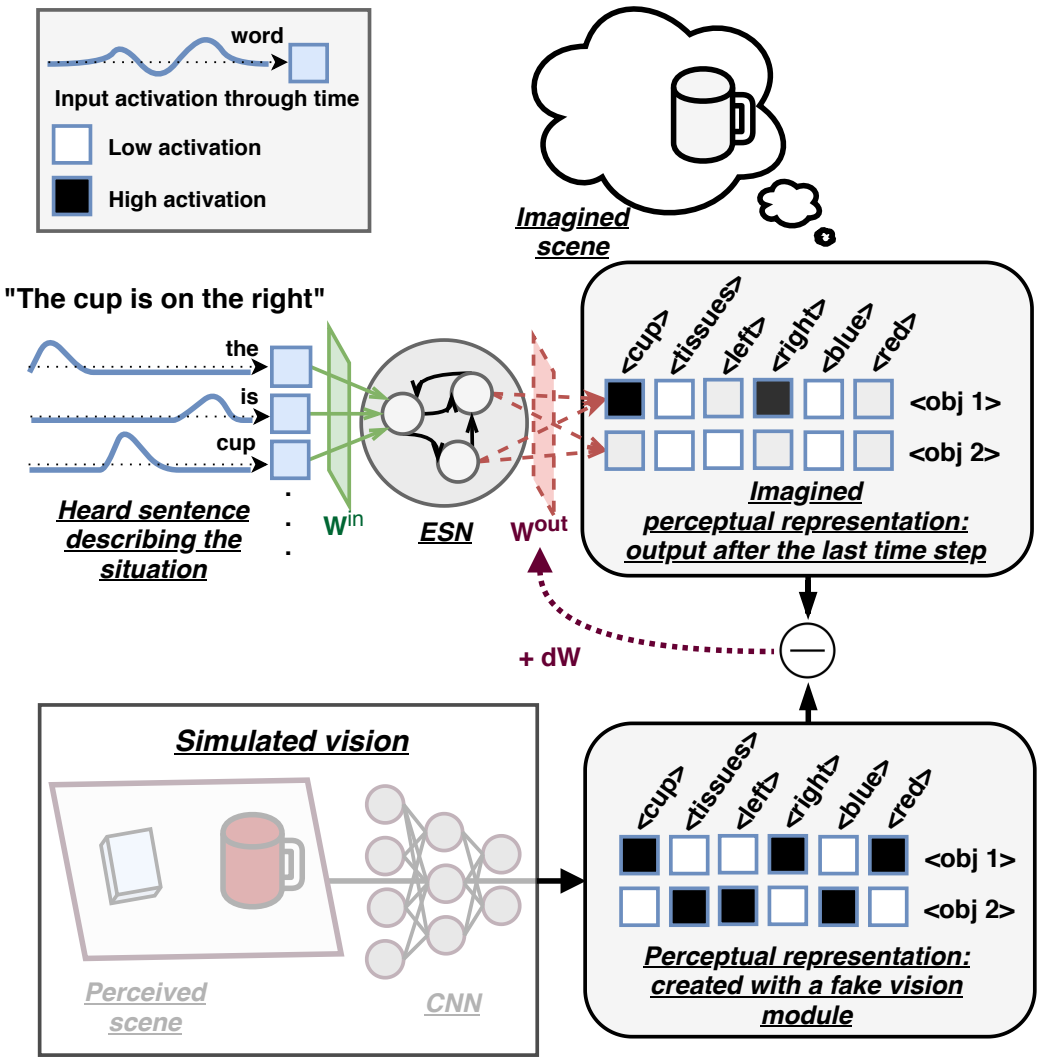

Figure 5.5: Training a reservoir language model in *Cross-Situational Learning* fashion. The teacher signal is assumed to be given by symbolic representations a robot can have from a scene with one or several objects. The difference between the network output and the vision module output is used to update slightly the network weights – using online FORCE learning [Sussillo & Abbott 2009] – at each time step or only at the last time step of the sentence. By repeating this step on various sentence-scene pairs, only the pertinent modifications should be kept, and the network would extract the semantics from the sequence of words in each sentence. Polysemous meaning of words can be learnt (e.g. "orange" color or object). Image from [Juven & Hinaut 2020]. This preliminary work was extended in several studies [Dinh & Hinaut 2020][Variengien & Hinaut 2020][Oota *et al.* 2022].

mous and synonymous words (e.g. "An orange orange is in the middle/center."). We tried different concept representations to enable generalization from one-object

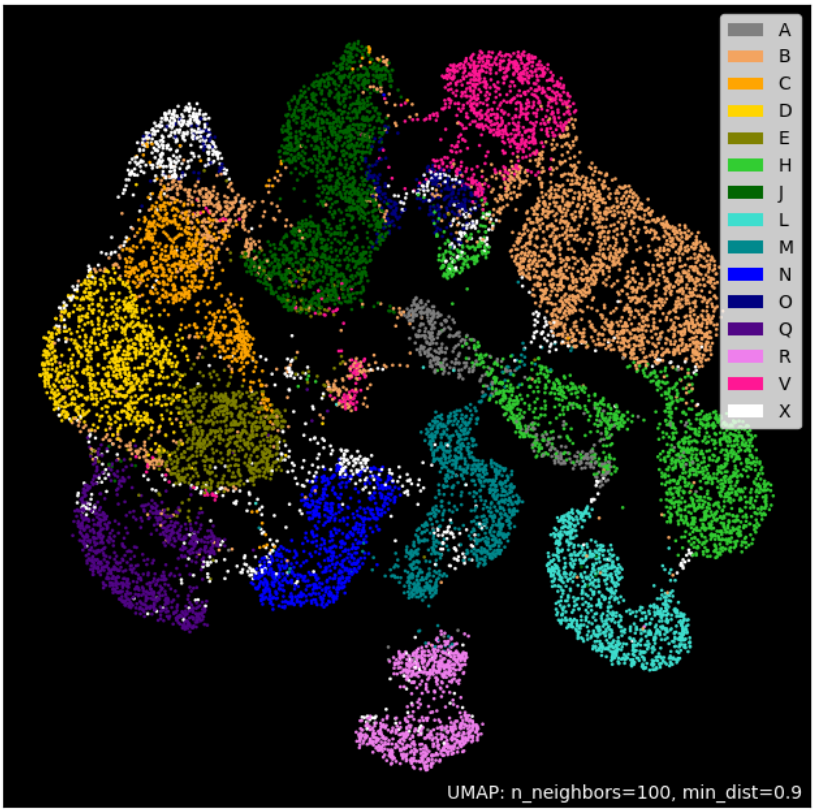

Figure 5.6: UMAP representation of 16k canary syllables generated from our Canary WaveGAN. Each cluster/color correspond to one class of the repertoire and class X (in white) represents the "unknown" class (i.e. outside existing classes). One can see the continuity between the syllables generated and the presence of interpolated classes (in white) which are not recognized as existing glasses. Syllable sounds generated where transformed to spectrogram before using UMAP [McInnes *et al.* 2018]. Image from [Pagliarini *et al.* 2021c].

sentences to two-object sentences [Dinh & Hinaut 2020], and we have preliminary results to ground such experiment with the MSCOCO dataset [Lin *et al.* 2014]. Finally, we demonstrated that reservoirs demonstrate better generalization than LSTMs when the vocabulary size increases [Variengien & Hinaut 2020] while the number of training sentences remain small (1000), and we compared generalization mechanisms of reservoirs and LSTMs with dimension reduction methods (UMAP) [McInnes *et al.* 2018].

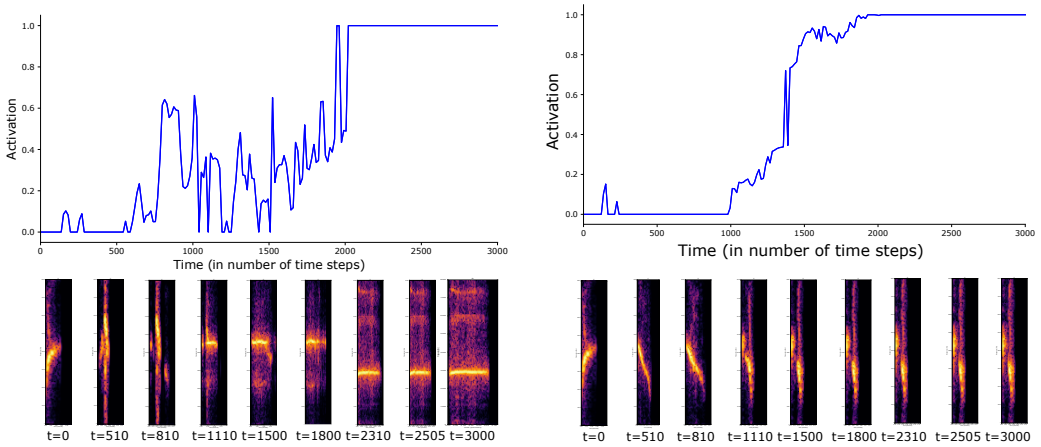

Figure 5.7: **Evolution of sound produced during learning for syllables *R* and *B1*.** (top) Evolution of the sensory response activation of unit R (left) and B1 (right) obtained during one instance of training. (bottom) Evolution of the corresponding sounds produced over time for nine selected time steps. Image from [Pagliarini *et al.* 2021a].

In a third batch of studies, we managed to train GANs for raw sounds (WaveGAN) with a large dataset of canary syllables (16000 renditions) and constrain the latent space to small dimensions (from 1 to 6) [Pagliarini *et al.* 2021c]. The sounds produced by the generators were identified and evaluated by a reservoir-based classifier trained on the same dataset. We also performed qualitative evaluation (using UMAP) of the GAN output spectrograms across GAN training epochs and latent dimensions. UMAP representations show the similarities between the training data and the generated data, and between the generated syllables and the interpolations produced. By exploring the latent representations of syllable types, we showed that they form well identifiable subspaces of the latent space, while producing some interpolations between syllables (see Figure 5.6). In another study [Pagliarini *et al.* 2021a], we used such low-dimensional GANs as motor control function to build a vocal sensorimotor model with the full action-perception loop with real sounds produced (see Figure 5.2). The sensory response function was modelled using a reservoir trained to decode canary syllables. We showed that a simple Hebbian learning rule, used for the inverse model, was able to learn the majority of the 16 canary syllables (see Figure 5.7). In the meanwhile, we showed that reservoirs generalize with less data then LSTMs for canary song la-

belling [Trouvain & Hinaut 2021].

5.1.4 Methodology and risk management

The project is organized around six work packages described bellow.

WP1: Hierarchy of Recurrent Neural Networks (RNNs) WP2: Speech sensori-motor model WP3: Action-Perception (AP) mechanism WP4: Grounding and Human-Robot Interaction (HRI) experiments with the Nao robot WP5: Dissemination and Crowdsourcing WP6: Management

5.1.4.1 WP1: Hierarchy of RNNs for sentence comprehension and production

Tasks: (1) To create a model learning to process and produce sentences; (2) Explore abstract sentence goal representations for sentence production; (3) Explore plausibility checking mechanism for sentence comprehension.

Preparatory work will be to continue current work on hierarchical reservoirs for sentence processing – i.e. Part-of-Speech (POS) and Semantic Role Labelling (SRL) – from speech (i.e. Mel-Frequency Cepstral Coefficients (MFCC)). We have already shown that such simple hierarchy already enables to correct errors in the bottom-up flow [Pedrelli & Hinaut 2020, Pedrelli & Hinaut 2022]. We will investigate how simple backwards connections (from top to bottom reservoir layers) could be added to enable to propagate such error correction ability at less abstract levels.

Concerning preparatory work, we will train the models using supervised learning in order to make a link with previously developed models [Hinaut & Dominey 2013, Hinaut *et al.* 2014]. We will use artificially generated data [Hinaut & Dominey 2013, Juven & Hinaut 2020] (useful for controlling parameters such as clause embedding and the complexity of the corpus in general), robot commands from human-robot interaction [Dukes 2014, Hinaut *et al.* 2014, Hinaut *et al.* 2015b, Hinaut & Twiefel 2020] and data from other studies aiming to model language acquisition [Connor *et al.* 2008]. This will include corpora in different languages [Hinaut *et al.* 2015b]. We will see how such model is able to represent the meaning of the sentences (by using classical sentence parsing, binding semantic roles to words, etc. and more robot concept-based representations) in a way that is both suitable for sentence processing and sentence production. From the production side, we will call such top high-level representation: "abstract sentence goal representations". We will start by training each intermediate layer with the same temporal teacher signals than for the processing mode. We will then experiment mixtures of resulting representations between processing and production modes. Concerning the representation of the sequence of words and morphemes that will be inputted to

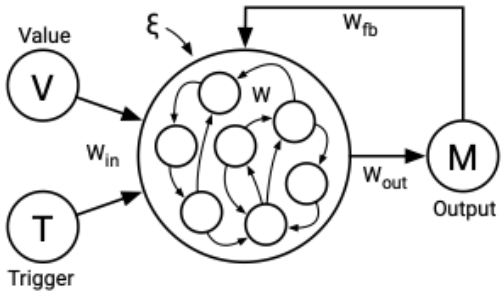

Figure 5.8: Robust model of gated working memory. It is an extension of a classical reservoir that we have made to extend the reservoir long-term memory abilities [Strock *et al.* 2020]. The read-out is called a *WM-unit*: it is trained to output the scalar value V when the trigger T is ON, and gate this value until the next trigger comes. Even if a stream of disturbing values is coming as input, the model is able to keep the gated value. Analysis of the reservoir shows that it behaves as a *linear attractor*. Internal neuronal states can show constant or dynamic neural activities – both observed in neurophysiology – while values are gated depending on parameters of the model.

and outputted by the model, we will experiment various representations in order to test their influence on the computations performed. For instance, we will consider arbitrary representations (orthogonal encoding), word embedding representations such as Word2Vec, BERT, etc.

The last task of this WP will be to explore mechanisms able to check this "viability of the interpretation" of the sentence, that we will call more simply "plausibility check" as shown in Figure 5.1. We will experiment various solutions implementing this plausibility check (e.g. based on adversarial methods such as GANs) and evaluate their influence on the processing and performances of the model. We will use fully supervised and weakly supervised trainings by using CSL [Juven & Hinaut 2020, Variengien & Hinaut 2020]. Afterwards, we will be able to connect the top of the hierarchy to a plausibility module and/or grounding module (link to WP4): this high-level module will evaluate the representation proposed by the network and either validate (i.e. sentence understood) or "ask for exploration" by sending a modulatory signal that will "flip" the symbols at points in the hierarchy where there is the ambiguity measure is the highest. This exploration mechanism will be repeated until it converges towards an evaluation of the plausibility that is satisfactory. The plausibility module does not need to be grounded as part of this WP, e.g. it can be a symbolic reasoning system able to state the plausibility of the predicates/SRL found, it will be grounded in WP4. The production mode will be bootstrapping from "abstract sentence goal representations": the model will first target easier goals (e.g. reproducing single words or expressions) before producing complete sentences. For sound production, we will rely on our findings on low-dimensional GAN generator model [Pagliarini *et al.* 2021c][Pagliarini *et al.* 2021a]

and apply it to speech (instead of birdsong).

Similarly, as in WP2, 3 and 4, we will compare the model to developmental language acquisition studies. Additionally, we will explore the effects of processing bilingual corpora. In particular, we will explore if the network produced mixed representations or tend to cluster representations in different languages. We will compare this to bilingual studies. In addition, we will explore how the model can process and produce *code-switched* sentences (i.e. sentences that have words from two languages) [Van Hell *et al.* 2015]. Interestingly, bilingual language production has been shown to involve cognitive control mechanisms [Rougier *et al.* 2005] in language switching studies: we will explore if such control mechanisms emerge in the architecture. This bilingual and code-switching experiments will be done in collaboration with S.L. Frank at Radboud University (Nijmegen, NL) with whom we started a collaboration thanks to a *Campus France VAN GOGH 2021* travel grant. With the help of a neurolinguist collaborator, Gaël Jobart also at the Bordeaux NeuroCampus, we will map internal activities, representations and supposed functions of the different parts of the developed models with brain activity of fMRI studies [Pallier *et al.* 2011, Nelson *et al.* 2017, Bhattasali *et al.* 2019]. For instance, we could use publicly available datasets – we already started with some corpora released by Nastase et al. [Nastase *et al.* 2021] – or on future protocols developed jointly. In particular, we will be interested to focus on multilingual datasets [Li *et al.* 2022] because of our previous works on reservoir models processing multiple languages [Hinaut *et al.* 2015b][Hinaut & Twiefel 2020], including code-switching [Detraz & Hinaut 2019a]. Such correspondences between model components and brain areas could suggest new hypotheses to be tested back in neuroimaging experiments; participating to a fruitful interacting loop between fMRI and computational experiments.

Risks. To minimize risks on the ability of models to perform the tasks we will compare the generalization performance from various RNN hierarchies (reservoir, LSTM [Hochreiter & Schmidhuber 1997], Gated Recurrent Unit (GRU) [Cho *et al.* 2014b, Cho *et al.* 2014a], reservoirs with hyperparameters optimized with BPTT, etc.). This will enable us to analyze the dynamics of the RNNs in order to compare how they generalize, like we did in [Variengien & Hinaut 2020] for simple non-hierarchical LSTMs and reservoirs. This will be of great use to the community, because it will enable to make a gradient of RNN hierarchies and to make our models easily comparable to the existing literature.

WP1 Expected outcome. This WP will provide first versions of the sentence comprehension and production model as general hierarchical RNNs. It will include plausibility check mechanism able to construct, by exploration, a plausible meaning representation of the processed sentence. This will be done by exploring potential ambiguities in speech (at various degrees of abstraction: e.g. syllables, words), a plausible meaning representation of the processed sentence. This WP will span from speech percepts (and sound production) to sentence meaning representations.

5.1.4.2 WP2: Speech sensorimotor model

Tasks: We will make successive vocal sensorimotor models learning (1) phonemes, (2) syllables, (3) morphemes and words.

In this WP we intend to develop vocal sensorimotor models learning to imitate human vocalizations by goal- exploration. We will build upon previous important works groups such as Schwartz, Bessière, Diard, Moulin-Frier, etc. [Schwartz *et al.* 2012, Moulin-Frier *et al.* 2015, Barnaud *et al.* 2018, Barnaud *et al.* 2019, Nabé *et al.* 2021, Nabé *et al.* 2022] for the "COSMO approach", on the one hand, and from Philippsen, Reinhart and Wrede [Philippsen *et al.* 2014, Philippsen *et al.* 2016, Philippsen 2021], on the other hand (see [Pagliarini *et al.* 2021b] for a review). We will rely on our experience with birdsong sensorimotor models generating real canary sounds with a low-dimensional GAN [Pagliarini *et al.* 2021c][Pagliarini *et al.* 2021a]. We will also use real corpora of children and human vocalizations based on data available online (e.g. HomeBank [VanDam *et al.* 2016] and CHILDES [MacWhinney 2014] projects).

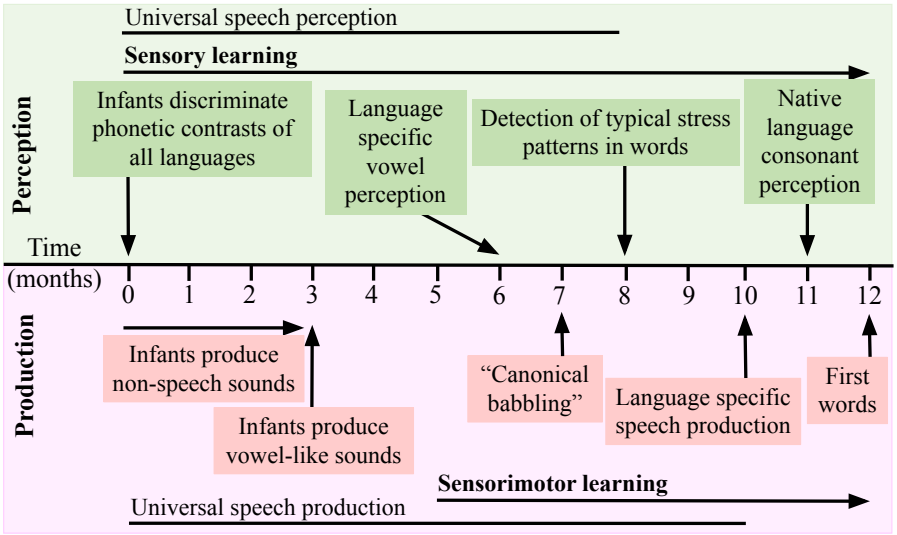

Figure 5.9: **First year of infant speech-perception and speech-production development.** Speech perception development (green background) is characterised by a sensory learning phase that shapes perception, from an initially universal perception to language-specific phoneme discrimination. Speech production development (pink background) is characterised by some preliminary phases followed by sensorimotor learning, where *canonical babbling* takes place. Image adapted from Doupe and Kuhl (1999) [Doupe & Kuhl 1999]. Image from [Pagliarini *et al.* 2021b].

First, a simple sensorimotor model imitating phonemes will be developed. The first conditions will bootstrap the experimentation with synthetized data [Philippsen 2021] and will then go towards real corpora. The model will learn to correctly represent perceptuo-motor gestures of phonemes. Secondly, we will explore how the model could produce syllables (i.e. composition of phonemes): we will test two conditions, one with only one layer sensorimotor model, and one with adding a supplementary sensorimotor model on top. Thus, we will experiment which architecture is best suitable for the formation of "syllable-like" perceptuo-motor gestures. Finally, we will explore what are the architectural conditions (i.e. how many sensorimotor layers) enabling the model to learn to imitate morphemes and words. Throughout these developments we will analyze the model behavior and compare it with developmental language studies (and other similar models) [Tomasello 2003]. We will test several models for the motor control function in order to imitate human vocal productions [Pagliarini *et al.* 2021b]: for instance Vocal-TractLab [Birkholz *et al.* 2006] and DIVA [Tourville & Guenther 2011]. Concerning the exploration of goals we will adapt previously developed ideas [Pagliarini *et al.* 2021b] of intrinsic motivation [Moulin-Frier *et al.* 2014] or goal-babbling [Rolf *et al.* 2010]. We will also use a low-dimensional GAN as motor control function for sound production, like we did in [Pagliarini *et al.* 2021c][Pagliarini *et al.* 2021a]. Our collaborator, Clément Moulin-Frier, also at Inria, will help with suggestions during the development of the speech sensorimotor models.

Risks. As we discussed in our review [Pagliarini *et al.* 2021b] such models may not be able to reproduce perfectly the distribution of sounds that could be obtained from real data, in particular, they may not be able to obtain a similar perceptuo-motor phase space than humans. Thus, we suggest that sensorimotor models should (ideally) learn only from syllables that they can correctly produce and perceive. Such condition may not be reached if such models learn from real human data (adult and children) vocalizations. Therefore, this is an additional reason to include two experimental conditions in our experiments, one with real data and one with purely simulated data (e.g. with vocal tract models).

WP2 Expected outcome. We will obtain a robust sensorimotor models based on coupled perceptuo-motor reservoirs that will be able to process stimuli (perception mode) or produce sounds (production mode) from various length (phonemes, morphemes, words). Stacked in few layers it will build more and more abstract representations of perceptuo-motor gestures.

5.1.4.3 WP3: Exploring Action-Perception (AP) mechanisms

Tasks: (1) Create the generic AP layer; (2) Explore conditions that let compositional symbols emerge; (3) Stack AP layers

Task 1: Create the generic AP layer. The idea of this first step is to define

the core layer of the architecture. In order to create the generic AP layer, we will explore various ways of combining the core functions the general sensorimotor architecture depicted in Figure 5.2. The goal is to explore how to make an active interaction between the perceptual space and the motor space, using inverse and forward models. We will mostly choose components based on random recurrent neural networks (i.e. reservoirs) with bio-plausible learning rules, in order to keep the overall architecture as homogeneous as possible and enable to intrinsically process continuous streams without unfolding time. Moreover, this will ease the analysis because the same tools could be used on various components. In this WP, we want to go beyond previous sensorimotor models [Pagliarini *et al.* 2021b] and have our AP layer actively recoding/shaping perceptual inputs or motor outputs. An AP layer could work in two "modes": a perceptive mode and a production mode.

Perceptive mode. The agent receives a sound stimulus (represented in acoustic or sensory space) which elicits a first representation in the perceptual space. Then, this activates a representation in the articulatory or motor space through the inverse model. Afterwards, activity is propagated again to the perceptual space through the forward model. The activities in the perceptual space are now mixed between outside acoustic stimuli and articulatory representations: the later will "refine" the perceptual representations in order to make them converge towards a stable perceptual category.

Production mode. First, a goal (i.e. a categorical perceptual representation) is activated in perceptual space. This produce an activation in motor space through the inverse model, which then produce a sound (in sensory space) through the motor control function. Then, this sound induces a new representation in perceptual space which lead to an adaptation of the motor command if the perceptual category perceived is different from the one triggered by the goal. Various variants of such hypothesized mechanism will be explored. For the learning framework we will explore various reinforcement learning like mechanisms [Warlaumont *et al.* 2013, Warlaumont & Finnegan 2016] adapted to recurrent neural networks: e.g. by means of "exploratory noise" with reward-modulated Hebbian learning [Hoerzer *et al.* 2014] or noise injected in the sensorimotor loop. At start, this WP focuses on building the core mechanisms, thus it will use rather simple sound input stimuli like sequences of pure tones, or canary syllables (that we already did in [Pagliarini *et al.* 2021c][Pagliarini *et al.* 2021a]), which could also be easily generated. Sound preprocessing will be performed with MFCCs (Mel Frequency Ceptrum Coefficents) like in our previous studies [Trouvain & Hinaut 2021, Pagliarini *et al.* 2021c][Pagliarini *et al.* 2021a].

Chunk continuous streams. The aim is to go towards models able to chunk continuous streams into discrete sequences of time-varying symbols. The stimuli will have to be chunked in "useful" pieces that will be processed by the production part and more abstract layers. Moreover, as Christiansen & Chater [Christiansen & Chater 2016, Christiansen *et al.* 2016] coin it in their [**?**] principle, stimuli not only have to be chunked, they have to be chunked as soon as possible, because new stimuli will arrive, and they will erase the current stimulus

if it is not processed quickly enough. This important time constrain is one reason which favors reservoir-based methods, because they are intrinsically processing time as such when they are trained, the learning algorithms are fast computationally and only need local information in time and space in order to learn. The task here will consist to find ways of quickly chunking stimuli without being overwhelmed by new inputs before the convergence of the categorical perception. We will thus explore methods enabling to chunk continuous streams, e.g. with populations of reservoirs self-supervising each other [Asabuki *et al.* 2018]. Besides the model developed by [McCauley & Christiansen 2019], although not based on neural networks, will probably be source of inspiration since it includes – along chunking – several features we want in the general model: incremental, online, local information, multilingual and works both for comprehension and production.

Task 2: Explore conditions that let compositional symbols emerge. We want to find mechanisms that will favor perceptual categories with compositional representation. An example of simple mechanism that can be implemented, to bootstrap the exploration, is to add a dynamic self-organizing map (DSOM) [Rougier & Boniface 2011] with several units that get activated (k-BMU, Best-Matching Units) instead of just one. This additional layer could be (1) added after the outputs (i.e. readout) units of each reservoir, or (2) replace the output layer [Pitti *et al.* 2020]. We assume this can be learned with methods such as 3-factor Hebbian learning rule with exploratory noise and other interesting features obtained with RNNs [Hoerzer *et al.* 2014, Pitti *et al.* 2022].

Task 3: Stack AP layers in a hierarchy. This task will explore how the previously developed AP layer (in tasks 1–3) could be stacked in few layers in order to bootstrap the way the hierarchy will work. The two main objectives of this task are to explore how the stacking of AP layers can lead to the transmission of "the right amount of information" (i) to the upper layers (bottom-up processes) and (ii) similarly to the lower layers (top-down processes). In this way, the information will flow to the upper layers in order to build more abstract information at each layer by considering bigger chunks of information; and conversely towards lower layers, in order to control the sub-goals that have to be performed until the precise motor commands that will be executed to produce sound.

Risks. In order to minimize the risks we will perform several models by incremental experiments. As a preliminary work (WP3's start), we will explore what are the conditions enabling the hierarchical embedding of symbols through developmental processes. We will make available tools needed for this WP directly inside the ReservoirPy library [Trouvain *et al.* 2020, Trouvain & Hinaut 2022, Trouvain *et al.* 2022], in order to make it easy and flexible to test various combinations of mechanisms. We proposed a general method for on-experts to optimise hyperparameters in a "non-blind" fashion in order to understand the relations between hyperparameters [Hinaut & Trouvain 2021]. In this preliminary work, we will explore a beta-test AP mechanism in a hierarchy (based on WP3) in order to have

first hints of the behavior in a hierarchy.

WP3 Expected outcome. We will obtain a robust Action-Perception (AP) mechanism based on coupled perceptuo-motor reservoirs that will be able to process a continuous stream of stimuli (perception mode) or produce a continuous stream of sounds (production mode). Stacked in few layers it will build more and more abstract representations of perceptuo-motor gestures.

5.1.4.4 WP4: Grounding experiments with the Nao robot

Tasks: (1) Integration of CSL trained models into robots and virtual agents; (2) Implement plausibility check mechanism in simulator and robot; (3) Integrate grounding as part of the hierarchy

We target to embody models into robots that will developmentally ground language from morphemes to sentences in order to better model how children acquire language and what could go wrong in developmental language disorders. We will start from preliminary results published [Dinh & Hinaut 2020, Juven & Hinaut 2020, Variengien & Hinaut 2020] and unpublished results on concept representations enabling to scale from one-object sentences to several-object sentences, in cross-situational learning conditions. The first task will be to extend our preliminary results with simulated vision [Dinh & Hinaut 2020, Juven & Hinaut 2020, Variengien & Hinaut 2020] with experiments on image datasets (e.g. MSCOCO [Lin *et al.* 2014]). Then, we will implement most robust models found in WP1–3 in a virtual environment (robotic simulation environments of iCub or Nao humanoid robots), before implementing them on the real Nao humanoid robot. The aim is to obtain symbolic representations that are a composition of goals and multimodal grounded representations. We will use a concrete corpus of sentences based on actions a robot can do, like in our previous studies [Hinaut *et al.* 2014, Hinaut & Twiefel 2020, Juven & Hinaut 2020] and current work [Oota *et al.* 2022]. We will base our studies on previously performed experiments with both sentence comprehension and production models with humanoid robots [Hinaut *et al.* 2014] such as Nao.

The grounding of the hierarchical network will be tested incrementally: we will "rebuild" the hierarchy with the grounding components step by step. Once bootstrapped on a first level of symbols (i.e. phonemes), further levels of abstraction will be added one by one, implementing goals that are more and more abstract, until quickly reaching the sentence level. The link with non-linguistic modalities will be performed with increasing levels of complexity. First, we will consider merging the representations from vision, with a pre-trained CNN (Convolutional Neural Network) such as AlexNet but with fewer parameters such as SqueezeNet [Iandola *et al.* 2016]. We will explore combinations of the last layers of such networks as representations forwarded to our hierarchical model. We will then consider proprioception modality which corresponds to the motor angles of

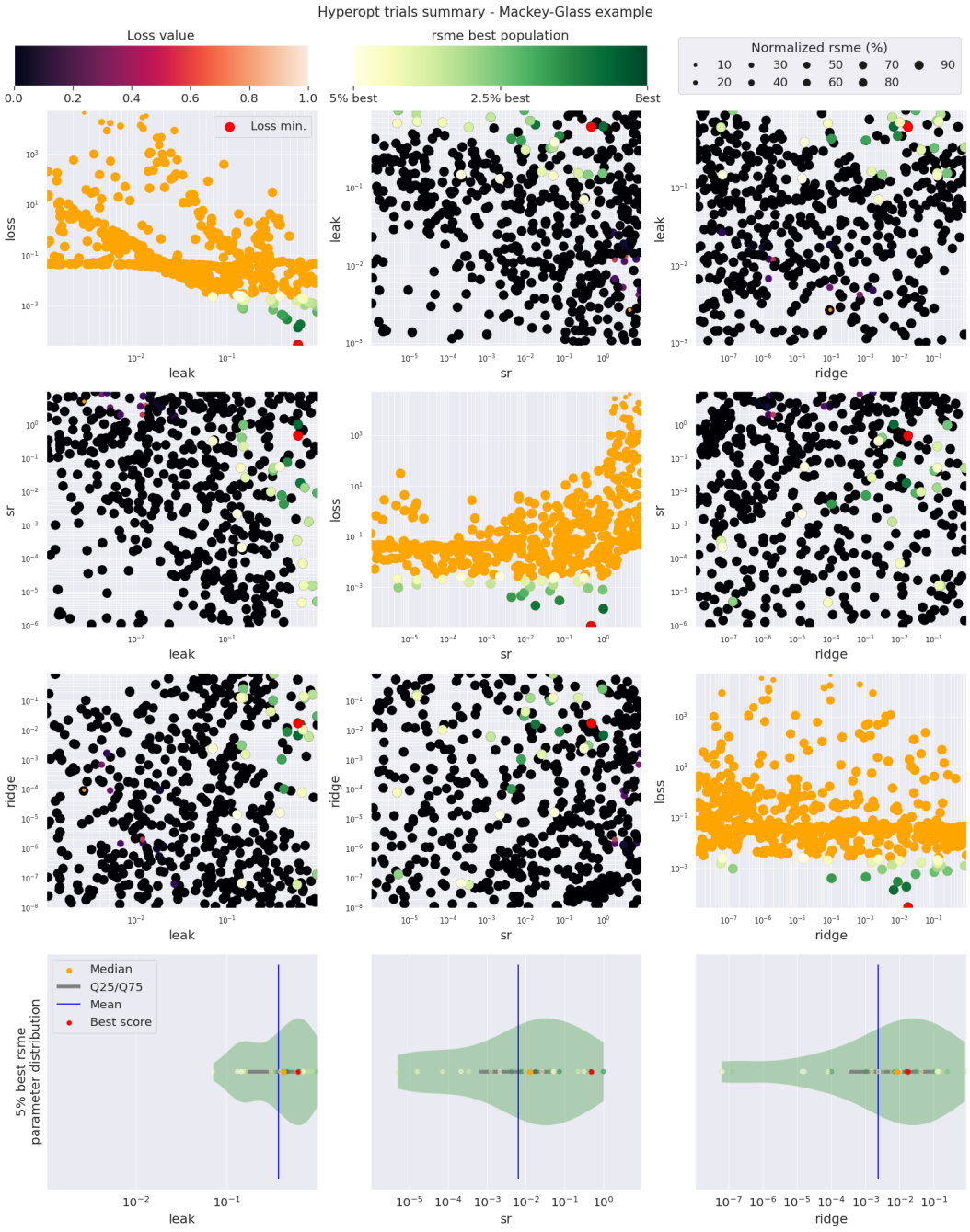

Figure 5.10: ReservoirPy graphical tool to explore optimal hyperparameters. Here, an example of figure obtained after a random search on 1000 trials to predict Mackey-Glass time series. The random search was performed on spectral radius (sr), leaking rate (leak) and regularization parameter (ridge). MSE and RMSE are displayed as evaluation metrics. Each trial point represent the averaged evaluation metrics over 10 sub-trials. Each sub-trial was performed on the same parameters combination within each trial, but with different ESN instances (e.g. different random weights initialization). Image from [Trouvain *et al.* 2020].

the various motors of the robot. We will get inspiration from various experiments on language grounding attempts with RNNs [Yamada *et al.* 2016]. In addition, we will implement a plausibility check mechanism that is best suited for robot simulators and for real robot experiments (see WP1). We will extensively study the effect of this grounding on the architecture and how the AP mechanism manages to generalize by integrated multimodal components in the hierarchy.

WP4 Expected outcome. This WP will provide a full hierarchical sentence model grounded in virtual and real robots. It will try to ground various levels of abstraction of perceptuo-motor gestures. The plausibility check mechanism will use the grounding in order to enhance the exploration of plausible meaning representations compared to non-grounded versions.

5.1.4.5 WP5: Crowdsourcing and Dissemination

Tasks: 1: We will (1) make a website to advertise the project and to collect data from online participants (crowdsourcing), (2) enhance our ReservoirPy library for internal and external use, and finally (3) disseminate demonstrations of models on the website and provide tools to run models online with one's own data.

The team will work on similar models that use common components (mostly reservoirs). To mitigate risks of incompatibility between developed models, we will actively ensure they compatibility through our ReservoirPy library (Figure 5.10 show an example of graphical tools provided). We will progressively enhance our library during the project, while the models are developed and based on it. Relying on a common library will enable us to disseminate our models more easily. In order to collect more data and at the same time make advertisement for the project, we will make a website collect data from online participants (crowdsourcing). We use preliminary work that I have done in collaboration with S. Wermter's lab and from various students work on what are the necessary mechanisms to make a good crowdsourcing website (e.g. by motivating people). Additionally, we will create a toolbox for internal and external use, in order to comply with our objectives of compatibility between models. Finally, we will disseminate demonstrations of models on the website and provide tools for non-programmers to test the models and even upload test with their own data.

WP5 Expected outcome. A library based on the models will enable efficiency in comparing and exchanging models within the team. This library shared with the scientific community will ease extensions of the models and application to various experimental data. The website will advertise the project and gather data thanks to crowdsourcing.

5.1.4.6 WP6: Management

This WP is a non-scientific WP dedicated to the management of the project.

Risks. Mitigation of risks are already described in each WP. Concerning the global timeline of the project, the WPs 1, 2 and 3 are organized temporally as to increase the complexity of the models, and thus the more exploratory architectures will be designed at last once we gained knowledge from the previous architectures. Making WP1, 2 and 3 independents – results from a WP are not needed for another one to start – in their development and organizing the WPs in this order permits to minimize the risks. With WP1 we will gain knowledge on different architectures of RNN hierarchies, which will be used when building hierarchies of sensorimotor models in WP2, which itself produce knowledge that will be used when building the hierarchy of AP layers in WP3. Whereas WP 4 and 5 are transversal throughout the project and will benefit from incremental improvements: I will make the continuity between the different steps and people involved by ensuring that documentation is well described. Both WPs will start early (or even before) the project: this will enable to narrow the scope of experiments with robots/humans if we encounter too important problems during preliminary experiments. Throughout the WPs, to evaluate and compare intermediate and high-level representations (with reservoirs or other compared RNNs), we will use quantitative and qualitative measures (separability of representations, UMAP, etc.). From the sensorimotor model, the aim is to imitate gestures (e.g. a particular word), thus if the model is able to correctly produce and perceptually cluster a syllable, it means it has correctly acquired this particular gesture. The variability in production and perception will be evaluated for each gesture. An indirect measure will be the ability of the representations to be both successfully used by bottom-up and top-down processes. Like in goal-directed exploration, the high-level representations used to produce a sentence should be found back (i.e. activate close representations) when the same sentence is processed. The same idea can be applied at different levels of hierarchies, even with incomplete hierarchies (e.g. from word level to concepts representations and back to word level). This ensures that evaluations can be performed at different stages of development, with different corpus complexity. The potential impact of COVID19 lockdowns will be limited, due to the fact that we get used in the team to work fully remotely for several months. The main impact could be for HRI experiments, limiting our ability to invite volunteers in the lab. However, this concerns only part of WP4 and we will also collect data via the crowdsourcing website. Moreover, during the last lockdown we were able to go at the lab if needed, thus we could still make HRI experiments with available people at the lab (as we are in a neuroscience lab, several people are present to continue experiments).

Ethical Issues. For all experiments involving humans (at the lab or online through a website), we will recruit adult and healthy participants. The inclusion criteria will be to speak the language in which we perform the experiment. Following participant information, they will fill up and sign an informed consent form indicating that (i) they can withdraw from the experiment at any time and (ii) the use of their data (text + audio + age range) should be restricted to the study or could be uploaded on a public repository for scientific data sharing. During experiments we will record written (text files) and/or spoken sentences (audio files) from the

participants. Sentences will be neutral and not related to any personal information. Datasets will include participants' recordings and age range. It will be anonymized prior to be used or shared. For human-robot interaction experiments, we will use the Nao robot which is light weight and could not physically harm users. Additional data will be obtained from other sources (linguistic corpora, fMRI, etc.) which are mainly anonymized public datasets, and, if it is not the case, we will anonymize them before use. Experiment details, data policy and informed consent forms will be evaluated by the COERLE (Inria Operational Committee for the assessment of Legal and Ethical risks), which is responsible for issuing an opinion on all requests for permission to conduct experimentation that is likely to affect the interests of people.

5.2 Insights on some related projects

Some Reservoir challenges

In order to build these models, we will have to tackle challenges with reservoirs: (1) build upon our Robust WM Reservoir Model [Strock *et al.* 2020] in order to incorporate gating mechanisms that can hold information to handle long-time dependencies like in GRU, LSTMs and other kinds of networks with gated units [Chung *et al.* 2014, Greff *et al.* 2016]; (2) make a reservoir extension model that is able to store *episodic memories* taking inspiration from hippocampus mechanisms [Chateau-Laurent & Alexandre 2021] such as pattern separation and pattern completion in high-dimensional space [Kassab & Alexandre 2018], and also taking inspiration from Neural Turing Machines (NTM) [Graves *et al.* 2014] and other memory-augmented networks [Rae *et al.* 2016]; (3) study how attention-like mechanisms [Vaswani *et al.* 2017] could be implemented in reservoirs; and (4) extend our previous studies [Juven & Hinaut 2020, Variengien & Hinaut 2020, Oota *et al.* 2022] to explore how reservoir generalizes on small and big corpora compared to other approaches like LSTMs and Transformers.

Some recent results on RNNs [Gumbsch *et al.* 2021] – and their use in hierarchical framework [Gumbsch *et al.* 2022] – from the groups of Bütz and Martius and may be useful as chunking mechanism: by applying a specific regularization on a gated kind of unit like GRU, it enables to obtain internal RNN representations that segment events. This work is part of the line of research followed by Bütz and colleagues on *Event-Predictive Cognition* [Butz *et al.* 2020], in relation with the *Theory of Event Coding* of Hommel and colleagues [Hommel *et al.* 2001] – proposing that actions and their effects are compressed into a common code – and the *Event segmentation theory* from Zacks and colleagues [Zacks *et al.* 2007] – suggesting that events are encoded, perceived, and processed as integrated units of thought.

Recently, Flynn et al. [Flynn *et al.* 2021] proposed a method to enable multifunctionality in a single reservoir: they train it to reproduce the climate (qualitatively similar dynamical behavior) of two different chaotic attractors. This is interesting in order to generate different kinds of dynamics from a single RNN (e.g. different

kinds of motor behaviors). As a by-product of their training method, they obtained *untrained attractors* which are different from the two different chaotic attractors they train their reservoir on. They explore how these untrained attractors appear and disappear studying the phase space of the spectral radius and they blending training hyperparameter – which reminds the *aperture* parameter used by Jaeger in his *Conceptors* [Jaeger 2014][Jaeger 2017]. Similarly, it would be interesting to investigate the existence of such *untrained attractors* in our Robust Working Memory (WM) Reservoir Model [Strock *et al.* 2020] when learning several line-attractors (each of them linked to one continuous WM-unit). We also developed Conceptors of such WM reservoir model, studying the effect of interpolating conceptors with different values stored in working memory [Strock *et al.* 2022]. It would be interesting to see (1) how these untrained attractor relate to Conceptors, e.g. if they could be interpolated or trained as Conceptors, and (2) if training method could be used to better interpolated between saved WM values.

The discovery of new features in hierarchical reservoirs is interesting in itself as it can provide interesting properties. For instance, deep ESNs provide multiple time-scale and an increase of richness of the dynamics [Gallicchio *et al.* 2017]. Moreover, hierarchical spiking models with topographic connectivity has been shown to improve computational performance and act as signal denoising [Zajzon *et al.* 2018, Zajzon *et al.* 2022].

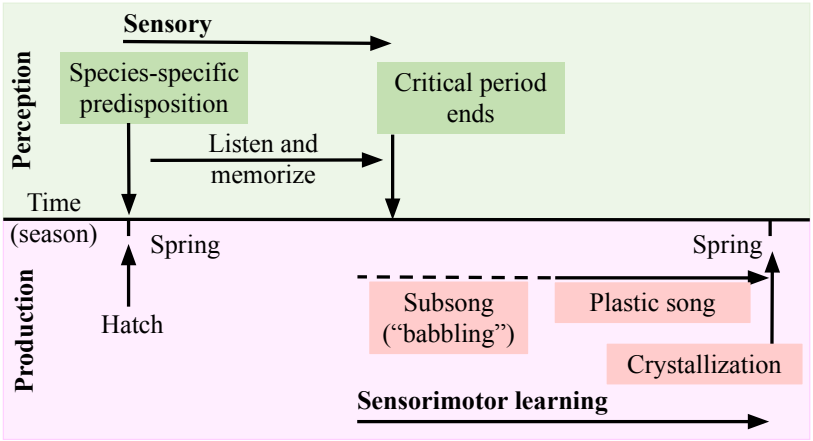

Figure 5.11: Imitative learning phases in birds. Three main phases characterise imitative learning in songbirds: the sensory learning phase, the sensorimotor learning phase (starting with subsong and continuing with a plastic song), and crystallization of the song (i.e. convergence to adult song). Image adapted from Doupe and Kuhl (1999) [Doupe & Kuhl 1999]. Image from [Pagliarini *et al.* 2021b].

In a more general fashion than a single hierarchy of reservoirs, exploring more distributed approaches could give important insights in the distributed nature of computations the brain makes. For instance, (1) exploring how a Recursive Self-Organizing Map (RecSOM) [Voegtlin 2002] composed of reservoirs, empowered with reinforcement learning, could automatically adapt to different kinds of subtasks; or (2) exploring how to design more complex architectures of reservoirs through evolutionary learning (such as some studies that proposed it for deep learning [Miikkulainen *et al.* 2019]).

Songbird sensorimotor models

Following Silvia Pagliarini's PhD, similarly to DeepPool project the aim is to build a hierarchical model that is able to learn process and produce from simple syllables to full songs. This would enable a full interactive loop with the environment (see Figure 5.11). We could then explore the possibility to transfer this hierarchical model to a continuous dynamical system, for instance using physical syrinx models [Amador *et al.* 2013] : even if the sounds produced would not be as good as what we obtained with a GAN [Pagliarini *et al.* 2021c], it would be interesting to explore what this hierarchical dynamical system allows. Such model could then be used to perform decoding with electrophysiological data from the data recorded in previous project on canaries in collaboration with Catherine Del Negro (see next paragraph). As we were able to obtain a high quality canary syllable production with a GAN generator during Silvia's PhD [Pagliarini *et al.* 2021c], this could open new experimental project with direct interaction between the models and the songbirds.

Birdsong analyses and electrophysiological experiments

On another topic related to sequences of symbols, chunk, syntax, and songbirds, I started a collaboration during my postdoc with Catherine Del Negro and Aurore Cazala on canary song analysis and neurophysiological experiments. Figure 5.12 shows an experimental protocol that we did with canaries. Some works were presented at conferences [Hinaut *et al.* 2017] or included in the PhD thesis of Aurore Cazala ([Cazala 2019] see pp. 77–110), but several parts of these works are still unpublished work. These studies included building tools to analyse automatically canary songs done with Nathan Trouvain [Trouvain & Hinaut 2021] and release an open source canary dataset on Zenodo [Giraudon *et al.* 2021]. I will pursue this line of works, making some links with the results obtained on songbird sensorimotor model [Pagliarini *et al.* 2021a] and GAN syllable generation [Pagliarini *et al.* 2021c].

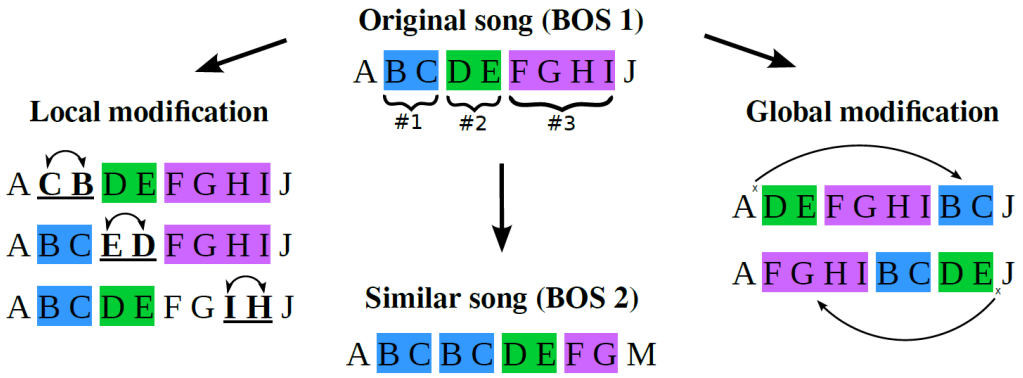

Figure 5.12: Experimental protocol to study how chunks are encoded. HVC (used as a proper name) is a sensorimotor area of canaries that is active both during the production and perception of songs. It is believed to control the sequence of syllables a canary produces. BOS is a given Bird Own Song: a song that the given canary produced. The protocol aims to study the changes produced in the HVC recorded neurons when local or global changes are made in the song. The BOS and artificially modified BOS versions are replayed to the canary while it is anesthetized, and some neurons of HVC are recorded. Each colored box is a *chunk*, i.e. a sequence of syllables that the canary produce often in a row and which have high transition probability between syllables. Local changes correspond to the scrambling of a chunk by swapping two syllables, aiming at producing low or zero probability transitions. Global changes are obtained by changing the location of a chunk in the song. The protocol have the interesting property of producing local and global modifications while keeping the lengths of the songs identical and keeping the same acoustic properties of syllables.

Chapter 6

Discussion

Contents

6.1 Conclusions

With the research program proposed I managed to bring together almost all the topics of my previous and current topic of research: reservoirs, hierarchical architectures, exploring new reservoir mechanisms such as Working-Memory units, chunking, symbol emergence and grounding, sequences of symbols, language processing and production, language acquisition, speech and audio processing and production, sensorimotor models, robotics, generic models of the cortex, interaction with the environment, dealing with noisy and ambiguous inputs, and link with experimental data.

As the architectures and the mechanisms developed we will be general and not be specific to language, they could be used also for other modalities or any kind of time series[1]: for instance sign languages or more generally to learn and produce complex sequence of actions. It could also be used for sensorimotor modelling in other species than humans. Indeed, having vocal sensorimotor models that would need minimal changes to work for humans or birds is appealing. As for human language modeling, I hope that songbirds models could also be linked to experimental data.

Finally, our ReservoirPy library will enable us to quickly develop prototypes with complex architectures, explore learning rules variants and more generally compare various features. In this regard, implementing as well various kinds of spiking neurons[2] would be interesting to compare the influence of topology, learning rules, etc. on the dynamics, computational power and generalisation of both spiking and rate-coded neurons. Besides, it could be an interesting tool to compare spike and rate minimal models[3] needed to explain experimental data. Hopefully, ReservoirPy

[1] Besides, Transformers are a good example of architecture that was first developed for language and then used for other kinds of inputs.

[2] For instance, using existing spiking network simulators as backends.

[3] In the sense of Ockham's razor.

will gather a community around such models and more generally gather the reservoir computing community on common tools.

Of course this seems an ambitious project, maybe too ambitious. However, the idea here is to start small, to build minimal subparts of the general model and then once a full version is developed look at what should be enhanced in the next round, and iterate like this. The aim is not so much to solve the whole problem(s) and sub-problems at once, but rather to have "tangible" scientific questions that support directions of research. In other words, the aim is to build iteratively global models: the first version of the model will have to make several assumptions and shortcuts in order to obtain a global hierarchy working. Then, step by step we will complexify the model to take into account more evidence from experiments. Even if each step may not be so meaningful taken in isolation, it will already provide minimal help to the community to see or think about things in a different way.

As I claimed earlier in this manuscript, body and environment are important in language, but do we really need to include them in all experiments? Does it matter at all that these language models are embedded in robots? This question can be difficult to answer, in particular when considering that most versions of the model will not require robots to function before some grounding to other modality would be needed. However, it is useful to show that these models are robust against environmental noise and can be executed in interaction with the robot in real time. Moreover, discrepancies observed between simulation and real world experiments can be informative. It may gives us new perspectives or new constraints to take into account, that we would not have thought of if the model would just have been be a "brain in a vat". Such *sim2real* problems appearing when moving from simulation to real robots has to be tested early on in the project in order to mitigate risks. Using robots in experiments requires more resources and time, thus finding the right balance is important.

6.2 Perspectives

Beyond the scope of the project, robot experiments would be needed to push forward the human-robot interaction experiments with dialogues, e.g. targeting to model interpersonal synergy [Fusaroli *et al.* 2014]. An appealing experiment would be to go prior the acquisition of spoken language, at the root of the emergence of symbols, when turn-taking starts to take place in the interactions from a child and his/her caregiver [Rączaszek-Leonardi *et al.* 2018]. This could serve to model transitions from grounded understanding of individual signs to the understanding symbolic relations. Studying symbol emergence [Taniguchi *et al.* 2016, Taniguchi *et al.* 2018] using language games with robots [Taniguchi *et al.* 2022] would also be a good way to take into account the environmental constraints in symbolic communication. For instance, blackbirds living in noisy cities tend to increase the pitch of song elements [Nemeth *et al.* 2013].

However, in order to have the control over all observable variables,

"presymbolic" or "subsymbolic" experiments could be done with reservoir models in order to create a continuity between dynamical systems and symbol emergence [Pattee & Rączaszek-Leonardi 2012, Rączaszek-Leonardi & Kelso 2008]. For instance, by using a simple reservoir-based agent navigating in a maze [Chaix-Eichel *et al.* 2022] where decisions of turning left or right have to be made continuously. Studying the influence of the presence or absence of reservoir feedback connections, as well as studying the changes occurring while increasing such feedback connections. Moreover, it would be interesting to see if reservoir gated Working Memory (WM) models [Strock *et al.* 2020] could learn to gate decisions in their WM-units through reinforcement learning based rules [Hoerzer *et al.* 2014] and analyse if these decisions are encoded continuously or as symbolic values. Finally, studying the coupling dynamics between such interacting agents could lead to another set of experiments. For instance, extending WM-units to store oscillators instead of fixed values could enable these agents to synchronise their behaviors by synchronizing their oscillators.

6.3 Is backprop our future?

Several people in machine learning want to find some proof that the principle of backpropagation, or recent derivatives, are biologically plausible. Similarly, claiming that reservoir computing is biologically plausible may seem strange given the random reservoir weights that are often kept fixed.[4] In my view, backprop principles[5] come more from maths and physics while reservoir computing ones come more from biology or ecology. Of course this division is deliberatly caricatural, dynamical systems is a topic shared across these disciplines. By mastering maths and physics we were able to build rockets to put humans on the moon. But this does not mean that we should always engineer systems in this way: biology and evolution (and the associated stochastic processes) created very sophisticated animals like bats and octopuses. Even if we can not know *what is it like to be a bat*[6] [Nagel 1974], we recognized that these animals are well adapted and exploit intelligently the physical laws in a way we don't.

With backprop one tries to optimize all (or many) of the weights from a deep neural network, while reservoir computing exploits the properties of a given dynamical system as an agent could exploit the physical laws of the environment with its body. In other words, reservoir computing adapts to the existing dynamical constraints of a random recurrent network, while backprop tries to constraints the physics of neural network. Let's take a metaphor. You are on a beach and you want

[4] This randomness is probably partly responsible for the lack of interest of a part of the community compared to other approaches. But there is a fundamental difference of paradigms that makes me think that trying to cast backprop or similar alternatives as brain mechanisms can be the wrong track to follow.

[5] In particular when applied to many layers of neural networks or applied by virtualization of time in RNNs for BPTT.

[6] https://en.wikipedia.org/wiki/What_Is_It_Like_to_Be_a_Bat%3F

to find a good spot where there are good waves to surf. The "backprop way" would be to build an artificial island[7] to try to design the perfect beach spot where waves are as you want. The "reservoir way" would be to be patient and *observe*[8] where and when there are good spots to surf in function of the tide. Moreover, building an artificial beach is not a long-term solution, erosion and environmental changes will force you to redo it again and again. Online adaptation is a better strategy in the long run.

Exploiting available properties that "come for free" from the environment is much more efficient. If you have a body that enables you to make approximate movements you do not need to have a precise control. For instance, the pulp at your fingertips enables you to grasp a mug in an easier way than a robot with metallic hands could[9]. From a biological point of view, relying partially on randomness is less costly and "good enough" to solve many problems. Moreover, it makes the system more robust and flexible. Evolutionary methods such as genetic algorithms are efficient to find "good enough" solutions where classical methods can not in reasonable time. In robotics new paradigms emerged in the last decades. [Brooks 1986] proposed its subsomption architecture to have autonomous agents more reactive to their environment. Bodies of the robots was also part of these new paradigms: Pfeifer and others [Pfeifer & Bongard 2006, Pfeifer & Pitti 2012] proposed they had to be considered as part of their intelligence. However, these approaches are not enough considered given the fails that could be seen at some of the DARPA challenges on autonomous robotics [10]. Similarly, neural-based machine learning should not pursue blindly this "backprop way" and borrow more principles from reservoir computing and other paradigms.

[7] Ironically, some people do it.

[8] One recognizes the *read-out* idea of reservoir paradigm.

[9] Try to grasp objects with thimbles to see how difficult it is.

[10] https://www.discovermagazine.com/technology/the-most-epic-robot-fails-of-the-darpa-robotics-challenge

6.4 A thought experiment

What is it like to be (see Figure 6.1)?

Figure 6.1: "Spinalis", Theo Jasen. *Filum, 2021-heden (era of the brain).* FAB Festival 2022, Bordeaux. Image CC0 X. Hinaut, Garden of the Bordeaux Fine Art museum, 25 Sept. 2022

What is it like to be this one? [Nagel 1974]

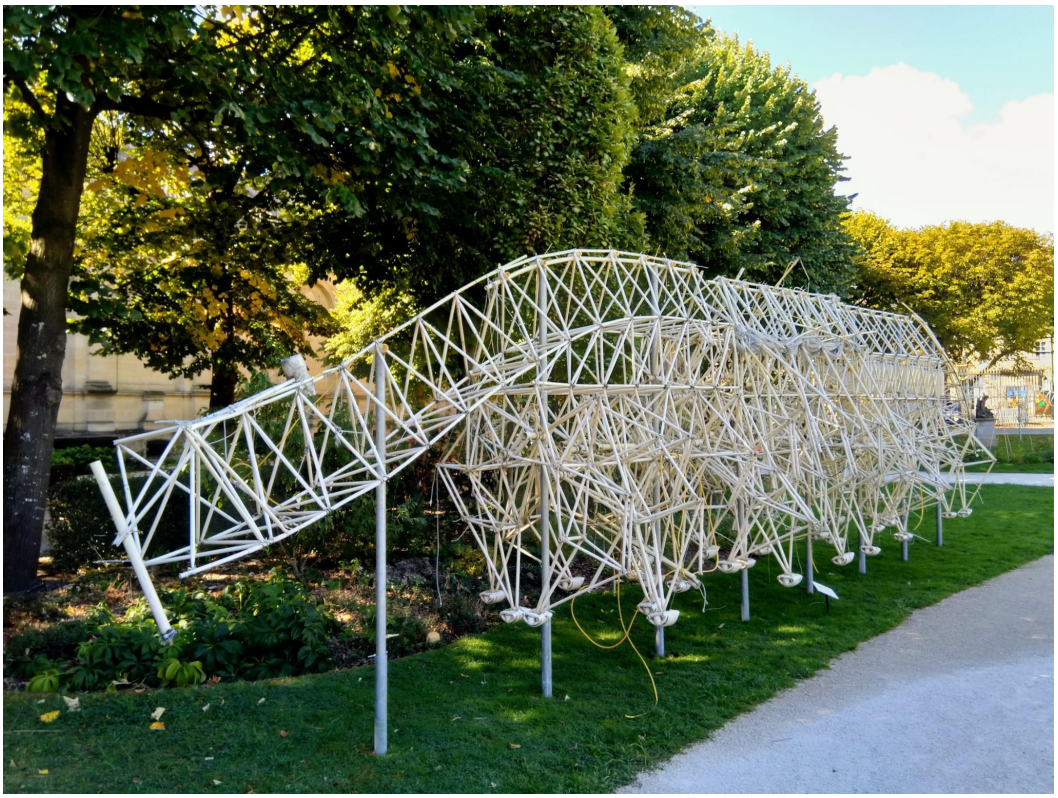

Figure 6.2: "Longus". Theo Jasen. *Cerebrum, 2006-2008 (era of the brain)*. FAB Festival 2022, Bordeaux. Image CC0 X. Hinaut, Garden of the Bordeaux Fine Art museum, 25 Sept. 2022

Imagine you see it moving.
Does your answer change?

As you probably recognize some artworks made of recycled tubes by Theo Jansen, you would probably answer "Nothing of course!" or maybe "Why do you even ask the question?". But if now you would have seen these artworks moving or if I would have shown you a video, would your answer would have been so quick? Maybe you would have tried for half a second to imagine yourself "being" that thing with your "mirror neuron system", maybe not to answer my question, but just to understand how this thing could "walk".

Now imagine that such a structure happens to walk on a very big sponge, how would its body react? Would it be able to walk over the sponge and get back to a normal walk after that? If it succeeds, would you consider it has having some coping skills, some intelligence?

Kevin O'Regan (see Figure 6.3) claims that *"Having the feeling of softness does not occur* in your brain*; rather, it resides in your noting that you are currently* interacting in a particular way with the sponge*."* [O'Regan 2011]. When reading this sentence out of its context we do not know if we need a brain to "feel softness", but for sure we need at least a body to interact with the sponge.

Figure 6.3: *"The feel of softness lies in the squishing. [...] Feeling softness is a quality of the* interaction *you have with a soft object like a sponge."* Text and image from [O'Regan 2011], pp.108–109.

LaMDA [Thoppilan *et al.* 2022] is one of the recent language model for dialog applications developed by Google. It is a family of Transformer-based neural language models specialized for dialog, which have up to 137 billion parameters and are pretrained on 1.56 trillion words of public dialog data and web text [Thoppilan *et al.* 2022]. Such numbers, $137 * 10^9$ weights and $1.56 * 10^{12}$ words, are huge; such number or words it is higher than what a human could say in a lifetime. In order to put this figures in perspective, let's make a very rough assumption that one neural network connection is equivalent to one of our biological

connection[11]. If we consider that we have 10^4 to 10^5 connections per neuron and about 10^{11} neurons in our brain [Herculano-Houzel 2009], this number can seem low compared to our 10^{15} to 10^{16} connections. However, one could wonder how much of these connections are actually used for language function, even if we include all embodied representations related to language.

Few months ago, some engineers at Google had an interaction with LaMDA and published their interaction[12]. It made one of them share its opinion on another blog post: talking about LaMDA, he does not understand why *"Google is resisting giving it what it wants"*[13]. Could this engineer answer the question "What is it like to be LaMDA?"

In an official blog post from Google, it is said that their *"systems still don't understand language the way people do"*, and just after it is said that *"many of our advanced models can understand information across languages or in non-language-based formats like images and videos"*[14]. I wonder how *softness* is represented in such "images and videos formats" as Kevin O'Regan tells us it lies in the interaction. Another question is whether *softness* representation depends on the language chosen [Regier & Kay 2009].

In your opinion, what is closer to our *feel of softness*? The one from the creatures of Theo Jansen, or the one from LaMDA? The question may seem strange, but I am not sure if everyone would answer the same. This is why it is interesting to ask. Now, suppose that you know that some of Jansen' creatures have a "neural system"[15] made "muscles and neurons", with some binary neurons able to perform logical operations such as NOT. This basic neural system enables them to detect the presence of water[16] in order to "run away" from the sea. Knowing that, would your answer (Theo's creature *vs.* LaMDA) change?

Such deep learning language models, although they are already used in many decoding fMRI experiments – and we started to use them also –, may not tell us how the mechanics of language acquisition and language processing unfolds in our brains. However, as they catch complex aspects of languages and dialogues they are a new kind of tool that should not be disregarded. Although, we should seriously take into account their environmental impact and advocate for transparency in this regard [Bender *et al.* 2021]. They also could be used indirectly. Given that they can have totally irrelevant answers which shows that they have no idea about what they are saying[17], they may be used as a *Reverse*

[11] Which I do not believe but that is not the point.

[12] *Is LaMDA Sentient? — an Interview*, Medium, June 11th 2022. https://cajundiscordian.medium.com/is-lamda-sentient-an-interview-ea64d916d917

[13] *What is LaMDA and what does it want?*, Medium, June 11th 2022. https://cajundiscordian.medium.com/what-is-lamda-and-what-does-it-want-688632134489

[14] *Understanding the world through language*, Google blog, May 11th 2022. https://blog.google/technology/ai/understanding-the-world-through-language/

[15] https://www.youtube.com/watch?v=75Z7-gmd_qk

[16] https://www.youtube.com/watch?v=JudNpGB07Jw

[17] Excerpt of exchanges reported in [Sejnowski 2022]: "D&D: When was Egypt transported for the second time across the Golden Gate Bridge? GPT-3: Egypt was transported for the second

Turing Test as they could be a mirror that reflects the intelligence of the interviewer [Sejnowski 2022]. Taking inspiration from works on coupling and synergies [Pfeifer & Pitti 2012, Pattee & Rączaszek-Leonardi 2012, Fusaroli *et al.* 2014], instead of doing large language models with this "backprop paradigm", why don't we aim for large language models that are, at the root, dynamical systems *coupled* with their environment and in *synergy* with other agents?

It is interesting to note that like evolution [Pfeifer & Pitti 2012], Theo Jansen did not include a neural system since the start in his creatures, he did it about 10 years after the first ones. [Pfeifer & Pitti 2012] propose that we have to *"adopt a continuous point of view of intelligence instead of considering that an agent is intelligent or not, because the exploitation of physics laws can be done at various degrees more of less important."*[18] With this continuous idea of intelligence we can imagine ourselves somehow connected to Theo Jansen' creatures with their "primitive brain" and their ability to walk around. Would you consider the *passive dynamic walking* legs of McGeer [McGeer *et al.* 1990] to be less intelligent than Jansen' ones because they seem less complex? From the perspective of [Pfeifer & Pitti 2012], without a body exploiting the laws of physics how could we find a continuity between LaMDA and us? Would you feel closer to Jansen' creature or to LaMDA?

time across the Golden Gate Bridge on October 13, 2017."

[18] My translation from French.

Bibliography

[Amador *et al.* 2013] A Amador, YS Perl, GB Mindlin and D Margoliash. *Elemental gesture dynamics are encoded by song premotor cortical neurons.* Nature, vol. 495, no. 7439, page 59, 2013. (Cited on page 65.)

[Asabuki *et al.* 2018] Toshitake Asabuki, Naoki Hiratani and Tomoki Fukai. *Interactive reservoir computing for chunking information streams.* PLOS Computational Biology, vol. 14, no. 10, page e1006400, 2018. (Cited on page 58.)

[Barnaud *et al.* 2018] Marie-Lou Barnaud, Pierre Bessière, Julien Diard and Jean-Luc Schwartz. *Reanalyzing neurocognitive data on the role of the motor system in speech perception within COSMO, a Bayesian perceptuo-motor model of speech communication.* Brain and language, vol. 187, pages 19–32, 2018. (Cited on page 55.)

[Barnaud *et al.* 2019] ML Barnaud, JL Schwartz, P Bessière and J Diard. *Computer simulations of coupled idiosyncrasies in speech perception and speech production with COSMO, a perceptuo-motor Bayesian model of speech communication.* PloS one, vol. 14, no. 1, page e0210302, 2019. (Cited on page 55.)

[Barone & Joseph 1989] P Barone and J-P Joseph. *Prefrontal cortex and spatial sequencing in macaque monkey.* Experimental brain research, vol. 78, no. 3, pages 447–464, 1989. (Cited on pages 19 and 20.)

[Barsalou 2008] Lawrence W. Barsalou. *Grounded Cognition.* Annual Review of Psychology, vol. 59, no. 1, pages 617–645, January 2008. (Cited on page 42.)

[beim Graben & Hutt 2015] Peter beim Graben and Axel Hutt. *Detecting event-related recurrences by symbolic analysis: applications to human language processing.* Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, vol. 373, no. 2034, page 20140089, 2015. (Cited on page 39.)

[Bender *et al.* 2021] Emily M Bender, Timnit Gebru, Angelina McMillan-Major and Shmargaret Shmitchell. *On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?.* In Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, pages 610–623, 2021. (Cited on page 74.)

[Bengio *et al.* 1994] Yoshua Bengio, Patrice Simard and Paolo Frasconi. *Learning long-term dependencies with gradient descent is difficult.* IEEE transactions on neural networks, vol. 5, no. 2, pages 157–166, 1994. (Cited on page 18.)

[Bhattasali *et al.* 2019] Shohini Bhattasali, Murielle Fabre, Wen-Ming Luh, Hazem Al Saied, Mathieu Constant, Christophe Pallier, Jonathan R. Brennan, R. Nathan Spreng and John Hale. *Localising memory retrieval and syntactic composition: an fMRI study of naturalistic language comprehension.* Language, Cognition and Neuroscience, vol. 34, no. 4, pages 491–510, 2019. (Cited on pages 42 and 54.)

[Birkholz *et al.* 2006] P Birkholz, D Jackèl and BJ Kroger. *Construction and control of a three-dimensional vocal tract model.* In ICASSP, volume 1. IEEE, 2006. (Cited on pages 44 and 56.)

[Brooks 1986] Rodney Brooks. *A robust layered control system for a mobile robot.* IEEE journal on robotics and automation, vol. 2, no. 1, pages 14–23, 1986. (Cited on page 70.)

[Brouwer & Hoeks 2013] Harm Brouwer and John C. J. Hoeks. *A time and place for language comprehension: mapping the N400 and the P600 to a minimal cortical network.* Frontiers in Human Neuroscience, vol. 7, 2013. (Cited on pages 39 and 44.)

[Brouwer *et al.* 2017] Harm Brouwer, Matthew W Crocker, Noortje J Venhuizen and John CJ Hoeks. *A neurocomputational model of the N400 and the P600 in language processing.* Cognitive science, vol. 41, pages 1318–1352, 2017. (Cited on pages 39, 40 and 44.)

[Buonomano & Merzenich 1995] Dean V Buonomano and Michael M Merzenich. *Temporal information transformed into a spatial code by a neural network with realistic properties.* Science, vol. 267, no. 5200, pages 1028–1030, 1995. (Cited on pages 18 and 19.)

[Butz *et al.* 2020] Martin V. Butz, Asya Achimova, David Bilkey and Alistair Knott. *Event-Predictive Cognition: A Root for Conceptual Human Thought.* Topics in Cognitive Science, vol. 13, no. 1, pages 10–24, December 2020. (Cited on page 63.)

[Cangelosi *et al.* 2010] Angelo Cangelosi, Giorgio Metta, Gerhard Sagerer, Stefano Nolfi, Chrystopher Nehaniv, Kerstin Fischer, Jun Tani, Tony Belpaeme, Giulio Sandini, Francesco Nori, Luciano Fadiga, Britta Wrede, Katharina Rohlfing, Elio Tuci, Kerstin Dautenhahn, Joe Saunders and Arne Zeschel. *Integration of Action and Language Knowledge: A Roadmap for Developmental Robotics.* IEEE Transactions on Autonomous Mental Development, vol. 2, no. 3, pages 167–195, September 2010. (Cited on page 45.)

[Caucheteux & King 2021] Charlotte Caucheteux and Jean-Rémi King. *Language processing in brains and deep neural networks: computational convergence and its limits.* BioRxiv, pages 2020–07, 2021. (Cited on page 43.)

[Cazala 2019] Aurore Cazala. *Codage neuronal de l'ordre des signaux acoustiques dans les chants des oiseaux*. PhD thesis, Université Paris Saclay (COmUE), 2019. (Cited on pages 3 and 65.)

[Chaix-Eichel *et al.* 2022] Naomi Chaix-Eichel, Snigdha Dagar, Quentin Lanneau, Karen Sobriel, Thomas Boraud, Frédéric Alexandre and Nicolas P Rougier. *From implicit learning to explicit representations*. arXiv preprint arXiv:2204.02484, 2022. (Cited on page 69.)

[Chang 2002] Franklin Chang. *Symbolically speaking: A connectionist model of sentence production*. Cognitive science, vol. 26, no. 5, pages 609–651, 2002. (Cited on page 43.)

[Chateau-Laurent & Alexandre 2021] Hugo Chateau-Laurent and Frédéric Alexandre. *Augmenting Machine Learning with Flexible Episodic Memory*. In 13th International Joint Conference on Computational Intelligence, 2021. (Cited on page 63.)

[Cho *et al.* 2014a] Kyunghyun Cho, Bart van Merriënboer, Dzmitry Bahdanau and Yoshua Bengio. *On the Properties of Neural Machine Translation: Encoder–Decoder Approaches*. In Proceedings of SSST-8, Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation, pages 103–111, 2014. (Cited on page 54.)

[Cho *et al.* 2014b] Kyunghyun Cho, Bart van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk and Yoshua Bengio. *Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation*. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 1724–1734, Doha, Qatar, October 2014. Association for Computational Linguistics. (Cited on pages 43 and 54.)

[Christiansen & Chater 2016] Morten H Christiansen and Nick Chater. *The now-or-never bottleneck: A fundamental constraint on language*. Behavioral and brain sciences, vol. 39, 2016. (Cited on pages 39 and 57.)

[Christiansen *et al.* 2016] Morten H Christiansen, Nick Chater and Peter W Culicover. Creating language: Integrating evolution, acquisition, and processing. MIT Press, 2016. (Cited on pages iii, 39, 40 and 57.)

[Chung *et al.* 2014] Junyoung Chung, Caglar Gulcehre, KyungHyun Cho and Yoshua Bengio. *Empirical evaluation of gated recurrent neural networks on sequence modeling*. arXiv preprint arXiv:1412.3555, 2014. (Cited on page 63.)

[Chung *et al.* 2016] Junyoung Chung, Sungjin Ahn and Yoshua Bengio. *Hierarchical multiscale recurrent neural networks*. arXiv preprint arXiv:1609.01704, 2016. (Cited on page 43.)

[Connor *et al.* 2008] Michael Connor, Yael Gertner, Cynthia Fisher and Dan Roth. *Baby srl: Modeling early language acquisition.* In Proceedings of the Twelfth Conference on Computational Natural Language Learning, pages 81–88. Association for Computational Linguistics, 2008. (Cited on page 52.)

[Crocker *et al.* 2006] Matthew W Crocker, Martin Pickering and Charles Clifton. Architectures and mechanisms for language processing. Cambridge University Press, 2006. (Cited on page 43.)

[Detraz & Hinaut 2019a] Pauline Detraz and Xavier Hinaut. *A Reservoir Model for Intra-Sentential Code-Switching Comprehension in French and English.* In CogSci'19 - 41st Annual Meeting of the Cognitive Science Society, Montréal, Canada, July 2019. (Cited on page 54.)

[Detraz & Hinaut 2019b] Pauline Detraz and Xavier Hinaut. *A Reservoir Model for Intra-Sentential Code-Switching Comprehension in French and English.* In CogSci'19-41st Annual Meeting of the Cognitive Science Society, 2019. (Cited on page 2.)

[Devlin *et al.* 2018] Jacob Devlin, Ming-Wei Chang, Kenton Lee and Kristina Toutanova. *Bert: Pre-training of deep bidirectional transformers for language understanding.* arXiv preprint arXiv:1810.04805, 2018. (Cited on pages 39 and 43.)

[Dinh & Hinaut 2020] Thanh Trung Dinh and Xavier Hinaut. *Language Acquisition with Echo State Networks: Towards Unsupervised Learning.* In ICDL 2020 - IEEE International Conference on Development and Learning, Valparaiso / Virtual, Chile, October 2020. (Cited on pages 3, 25, 47, 49, 51 and 59.)

[Dominey & Boucher 2005] Peter Ford Dominey and Jean-David Boucher. *Developmental stages of perception and language acquisition in a perceptually grounded robot.* Cognitive Systems Research, vol. 6, no. 3, pages 243–259, September 2005. (Cited on page 45.)

[Dominey *et al.* 1995] Peter Dominey, Michael Arbib and Jean-Paul Joseph. *A model of corticostriatal plasticity for learning oculomotor associations and sequences.* Journal of cognitive neuroscience, vol. 7, no. 3, pages 311–336, 1995. (Cited on page 19.)

[Dominey 1995] Peter F Dominey. *Complex sensory-motor sequence learning based on recurrent state representation and reinforcement learning.* Biological cybernetics, vol. 73, no. 3, pages 265–274, 1995. (Cited on pages 18 and 19.)

[Doupe & Kuhl 1999] AJ Doupe and PK Kuhl. *Birdsong and human speech: common themes and mechanisms.* Annual review of neuroscience, vol. 22, no. 1, pages 567–631, 1999. (Cited on pages 3, 55 and 64.)

[Dukes 2014] K. Dukes. *SemEval-2014 Task 6: Supervised Semantic Parsing of Robotic Spatial Commands*. SemEval 2014, page 45, 2014. (Cited on page 52.)

[Elman 1990] Jeffrey L Elman. *Finding structure in time*. Cognitive science, vol. 14, no. 2, pages 179–211, 1990. (Cited on page 43.)

[Enel *et al.* 2016] Pierre Enel, Emmanuel Procyk, René Quilodran and Peter Ford Dominey. *Reservoir computing properties of neural dynamics in prefrontal cortex*. PLoS computational biology, vol. 12, no. 6, page e1004967, 2016. (Cited on page 19.)

[Fadiga *et al.* 2009] Luciano Fadiga, Laila Craighero and Alessandro D'Ausilio. *Broca's Area in Language, Action, and Music*. Annals of the New York Academy of Sciences, vol. 1169, no. 1, pages 448–458, July 2009. (Cited on page 2.)

[Felleman & Van Essen 1991] Daniel J Felleman and David C Van Essen. *Distributed hierarchical processing in the primate cerebral cortex*. Cerebral cortex (New York, NY: 1991), vol. 1, no. 1, pages 1–47, 1991. (Cited on page 2.)

[Flynn *et al.* 2021] Andrew Flynn, Vassilios A. Tsachouridis and Andreas Amann. *Multifunctionality in a reservoir computer*. Chaos: An Interdisciplinary Journal of Nonlinear Science, vol. 31, no. 1, page 013125, January 2021. (Cited on page 63.)

[Friston 2018] Karl Friston. *Does predictive coding have a future?* Nature neuroscience, vol. 21, no. 8, pages 1019–1021, 2018. (Cited on page 43.)

[Fusaroli *et al.* 2014] Riccardo Fusaroli, Joanna Rączaszek-Leonardi and Kristian Tylén. *Dialog as interpersonal synergy*. New Ideas in Psychology, vol. 32, pages 147–157, 2014. (Cited on pages 68 and 75.)

[Gallicchio *et al.* 2017] Claudio Gallicchio, Alessio Micheli and Luca Pedrelli. *Deep reservoir computing: A critical experimental analysis*. Neurocomputing, vol. 268, pages 87–99, 2017. (Cited on page 64.)

[Garagnani & Pulvermüller 2016] Max Garagnani and Friedemann Pulvermüller. *Conceptual grounding of language in action and perception: a neurocomputational model of the emergence of category specificity and semantic hubs*. European Journal of Neuroscience, vol. 43, no. 6, pages 721–737, 2016. (Cited on page 40.)

[Garagnani *et al.* 2008] Max Garagnani, Thomas Wennekers and Friedemann Pulvermüller. *A neuroanatomically grounded Hebbian-learning model of attention–language interactions in the human brain*. European Journal of Neuroscience, vol. 27, no. 2, pages 492–513, 2008. (Cited on page 40.)

[Gers *et al.* 2000] Felix A Gers, Jürgen Schmidhuber and Fred Cummins. *Learning to forget: Continual prediction with LSTM.* Neural computation, vol. 12, no. 10, pages 2451–2471, 2000. (Cited on page 19.)

[Giraudon *et al.* 2021] Juliette Giraudon, Nathan Trouvain, Aurore Cazala, Catherine Del Negro and Xavier Hinaut. *Labeled songs of domestic canary M1-2016-spring (Serinus canaria)*, 2021. (Cited on pages 3 and 65.)

[Graves *et al.* 2013] Alex Graves, Abdel-rahman Mohamed and Geoffrey Hinton. *Speech recognition with deep recurrent neural networks.* In 2013 IEEE international conference on acoustics, speech and signal processing, pages 6645–6649. Ieee, 2013. (Cited on page 43.)

[Graves *et al.* 2014] Alex Graves, Greg Wayne and Ivo Danihelka. *Neural turing machines.* arXiv preprint arXiv:1410.5401, 2014. (Cited on page 63.)

[Greff *et al.* 2016] Klaus Greff, Rupesh K Srivastava, Jan Koutník, Bas R Steunebrink and Jürgen Schmidhuber. *LSTM: A search space odyssey.* IEEE transactions on neural networks and learning systems, vol. 28, no. 10, pages 2222–2232, 2016. (Cited on page 63.)

[Gumbsch *et al.* 2021] Christian Gumbsch, Martin V Butz and Georg Martius. *Sparsely changing latent states for prediction and planning in partially observable domains.* Advances in Neural Information Processing Systems, vol. 34, pages 17518–17531, 2021. (Cited on page 63.)

[Gumbsch *et al.* 2022] Christian Gumbsch, Maurits Adam, Birgit Elsner, Georg Martius and Martin V Butz. *Developing hierarchical anticipations via neural network-based event segmentation.* arXiv preprint arXiv:2206.02042, 2022. (Cited on page 63.)

[Haeusler & Maass 2007] Stefan Haeusler and Wolfgang Maass. *A statistical analysis of information-processing properties of lamina-specific cortical microcircuit models.* Cerebral cortex, vol. 17, no. 1, pages 149–162, 2007. (Cited on page 19.)

[Hagoort 2005] Peter Hagoort. *On Broca, brain, and binding: a new framework.* Trends in Cognitive Sciences, vol. 9, no. 9, pages 416–423, September 2005. (Cited on page 1.)

[Hamzei *et al.* 2003] Farsin Hamzei, Michel Rijntjes, Christian Dettmers, Volkmar Glauche, Cornelius Weiller and Christian Büchel. *The human action recognition system and its relationship to Broca's area: an fMRI study.* NeuroImage, vol. 19, no. 3, pages 637–644, July 2003. (Cited on page 1.)

[Herculano-Houzel 2009] Suzana Herculano-Houzel. *The human brain in numbers: a linearly scaled-up primate brain.* Frontiers in human neuroscience, page 31, 2009. (Cited on page 74.)

[Hinaut & Dominey 2011] Xavier Hinaut and Peter Ford Dominey. *A three-layered model of primate prefrontal cortex encodes identity and abstract categorical structure of behavioral sequences.* Journal of Physiology-Paris, vol. 105, no. 1-3, pages 16–24, 2011. (Cited on page 2.)

[Hinaut & Dominey 2012] Xavier Hinaut and Peter F Dominey. *On-line processing of grammatical structure using reservoir computing.* In International Conference on Artificial Neural Networks, pages 596–603. Springer, 2012. (Cited on page 2.)

[Hinaut & Dominey 2013] Xavier Hinaut and Peter Ford Dominey. *Real-Time Parallel Processing of Grammatical Structure in the Fronto-Striatal System: A Recurrent Network Simulation Study Using Reservoir Computing.* PLoS ONE, vol. 8, no. 2, page e52946, February 2013. (Cited on pages 2, 22, 43, 47 and 52.)

[Hinaut & Trouvain 2021] Xavier Hinaut and Nathan Trouvain. *Which Hype for my New Task? Hints and Random Search for Reservoir Computing Hyperparameters.* In ICANN 2021 - 30th International Conference on Artificial Neural Networks, Bratislava, Slovakia, September 2021. (Cited on pages 23 and 58.)

[Hinaut & Twiefel 2020] Xavier Hinaut and Johannes Twiefel. *Teach Your Robot Your Language! Trainable Neural Parser for Modeling Human Sentence Processing: Examples for 15 Languages.* IEEE Transactions on Cognitive and Developmental Systems, vol. 12, no. 2, pages 179–188, June 2020. (Cited on pages 3, 52, 54 and 59.)

[Hinaut & Wermter 2014] Xavier Hinaut and Stefan Wermter. *An incremental approach to language acquisition: Thematic role assignment with echo state networks.* In International Conference on Artificial Neural Networks, pages 33–40. Springer, 2014. (Cited on page 2.)

[Hinaut *et al.* 2014] Xavier Hinaut, Maxime Petit, Gregoire Pointeau and Peter F. Dominey. *Exploring the acquisition and production of grammatical constructions through human-robot interaction with echo state networks.* Frontiers in Neurorobotics, vol. 8, May 2014. (Cited on pages 2, 52 and 59.)

[Hinaut *et al.* 2015a] Xavier Hinaut, Florian Lance, Colas Droin, Maxime Petit, Gregoire Pointeau and Peter Ford Dominey. *Corticostriatal response selection in sentence production: Insights from neural network simulation with reservoir computing.* Brain and language, vol. 150, pages 54–68, 2015. (Cited on page 2.)

[Hinaut *et al.* 2015b] Xavier Hinaut, Johannes Twiefel, Maxime Petit, Peter Dominey and Stefan Wermter. *A Recurrent Neural Network for Multiple Language Acquisition: Starting with English and French.* In Proceedings of the

NIPS Workshop on Cognitive Computation: Integrating Neural and Symbolic Approaches (CoCo 2015), Montreal, Canada, December 2015. (Cited on pages 2, 52 and 54.)

[Hinaut *et al.* 2015c] Xavier Hinaut, Johannes Twiefel, Marcelo Borghetti Soares, Pablo Barros, Luiza Mici and Stefan Wermter. *Humanoidly Speaking -How the Nao humanoid robot can learn the name of objects and interact with them through common speech.* In International Joint Conference on Artificial Neural Networks – IJCAI, Video Competition, Buenos Aires, Argentina, July 2015. (Cited on page 2.)

[Hinaut *et al.* 2016] Xavier Hinaut, Johannes Twiefel and Stefan Wermter. *Recurrent Neural Network for Syntax Learning with Flexible Predicates for Robotic Architectures.* In The Sixth Joint IEEE International Conference Developmental Learning and Epigenetic Robotics (ICDL-EPIROB), Cergy, France, September 2016. (Cited on page 2.)

[Hinaut *et al.* 2017] Xavier Hinaut, Aurore Cazala and Catherine del Negro. *Neural coding of variable song structure in the songbird.* In EBM 2017 - European Birdsong Meeting, page 1, Bordeaux, France, May 2017. (Cited on pages 3 and 65.)

[Hinaut 2018] Xavier Hinaut. *Which input abstraction is better for a robot syntax acquisition model? phonemes, words or grammatical constructions?* In 2018 Joint IEEE 8th International Conference on Development and Learning and Epigenetic Robotics (ICDL-EpiRob), pages 281–286. IEEE, 2018. (Cited on page 2.)

[Hochreiter & Schmidhuber 1997] Sepp Hochreiter and Jürgen Schmidhuber. *Long Short-Term Memory.* Neural Computation, vol. 9, no. 8, pages 1735–1780, November 1997. (Cited on pages 18, 19, 47 and 54.)

[Hoerzer *et al.* 2014] Gregor M. Hoerzer, Robert Legenstein and Wolfgang Maass. *Emergence of Complex Computational Structures From Chaotic Neural Networks Through Reward-Modulated Hebbian Learning.* Cerebral Cortex, vol. 24, no. 3, pages 677–690, March 2014. (Cited on pages 57, 58 and 69.)

[Hommel *et al.* 2001] Bernhard Hommel, Jochen Müsseler, Gisa Aschersleben and Wolfgang Prinz. *The Theory of Event Coding (TEC): A framework for perception and action planning.* Behavioral and Brain Sciences, vol. 24, no. 5, pages 849–878, October 2001. (Cited on page 63.)

[Iandola *et al.* 2016] Forrest N Iandola, Song Han, Matthew W Moskewicz, Khalid Ashraf, William J Dally and Kurt Keutzer. *SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and< 0.5 MB model size.* arXiv preprint arXiv:1602.07360, 2016. (Cited on page 59.)

[Jaeger & Haas 2004] Herbert Jaeger and Harald Haas. *Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication.* Science, vol. 304, no. 5667, pages 78–80, 2004. (Cited on pages 18, 19 and 43.)

[Jaeger *et al.* 2007] Herbert Jaeger, Mantas Lukoševičius, Dan Popovici and Udo Siewert. *Optimization and applications of echo state networks with leaky-integrator neurons.* Neural networks, vol. 20, no. 3, pages 335–352, 2007. (Cited on page 23.)

[Jaeger 2001] H Jaeger. *The "echo state" approach to analysing and training recurrent neural networks.* Bonn, Germany: GMD Technical Report, vol. 148, no. 34, 2001. (Cited on pages 18, 19 and 23.)

[Jaeger 2002] Herbert Jaeger. *Adaptive nonlinear system identification with echo state networks.* Advances in neural information processing systems, vol. 15, 2002. (Cited on page 19.)

[Jaeger 2007] Herbert Jaeger. *Echo state network.* Scholarpedia, vol. 2, no. 9, page 2330, 2007. (Cited on page 18.)

[Jaeger 2014] Herbert Jaeger. *Controlling recurrent neural networks by conceptors.* arXiv preprint arXiv:1403.3369, 2014. (Cited on page 64.)

[Jaeger 2017] Herbert Jaeger. *Using Conceptors to Manage Neural Long-Term Memories for Temporal Patterns.* Journal of Machine Learning Research, vol. 18, no. 13, pages 1–43, 2017. (Cited on page 64.)

[Juven & Hinaut 2020] Alexis Juven and Xavier Hinaut. *Cross-Situational Learning with Reservoir Computing for Language Acquisition Modelling.* In 2020 International Joint Conference on Neural Networks (IJCNN). IEEE, July 2020. (Cited on pages 3, 17, 25, 36, 47, 49, 52, 53, 59 and 63.)

[Kassab & Alexandre 2018] Randa Kassab and Frédéric Alexandre. *Pattern separation in the hippocampus: distinct circuits under different conditions.* Brain Structure and Function, vol. 223, no. 6, pages 2785–2808, 2018. (Cited on page 63.)

[Kelso *et al.* 1995] JAS Kelso, P Case, T Holroyd, E Horvath, J Rączaszek, B Tuller and M Ding. *Multistability and metastability in perceptual and brain dynamics.* In Ambiguity in mind and nature, pages 159–184. Springer, 1995. (Cited on page 41.)

[Koechlin & Jubault 2006] Etienne Koechlin and Thomas Jubault. *Broca's Area and the Hierarchical Organization of Human Behavior.* Neuron, vol. 50, no. 6, pages 963–974, June 2006. (Cited on page 2.)

[Kröger & Bekolay 2019] Bernd J Kröger and Trevor Bekolay. Neural modeling of speech processing and speech learning. Springer, 2019. (Cited on page 44.)

[Li *et al.* 2022] Jixing Li, Shohini Bhattasali, Shulin Zhang, Berta Franzluebbers, Wen-Ming Luh, R Nathan Spreng, Jonathan R Brennan, Yiming Yang, Christophe Pallier and John Hale. *Le Petit Prince multilingual naturalistic fMRI corpus.* Scientific Data, vol. 9, no. 1, pages 1–15, 2022. (Cited on page 54.)

[Lin *et al.* 2014] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár and C Lawrence Zitnick. *Microsoft coco: Common objects in context.* In European conference on computer vision, pages 740–755. Springer, 2014. (Cited on pages 51 and 59.)

[Liu *et al.* 2019] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer and Veselin Stoyanov. *Roberta: A robustly optimized bert pretraining approach.* arXiv preprint arXiv:1907.11692, 2019. (Cited on page 43.)

[Lukoševičius & Jaeger 2009] Mantas Lukoševičius and Herbert Jaeger. *Reservoir computing approaches to recurrent neural network training.* Computer Science Review, vol. 3, no. 3, pages 127–149, 2009. (Cited on page 43.)

[Luong *et al.* 2015] Minh-Thang Luong, Hieu Pham and Christopher D Manning. *Effective approaches to attention-based neural machine translation.* arXiv preprint arXiv:1508.04025, 2015. (Cited on page 43.)

[Maass *et al.* 2002] Wolfgang Maass, Thomas Natschläger and Henry Markram. *Real-time computing without stable states: A new framework for neural computation based on perturbations.* Neural computation, vol. 14, no. 11, pages 2531–2560, 2002. (Cited on pages 18 and 19.)

[Machens *et al.* 2010] Christian K Machens, Ranulfo Romo and Carlos D Brody. *Functional, but not anatomical, separation of "what" and "when" in prefrontal cortex.* Journal of Neuroscience, vol. 30, no. 1, pages 350–360, 2010. (Cited on page 19.)

[MacWhinney 2014] Brian MacWhinney. The childes project: Tools for analyzing talk, volume ii: The database. Psychology Press, 2014. (Cited on page 55.)

[Markov & Kennedy 2013] Nikola T Markov and Henry Kennedy. *The importance of being hierarchical.* Current opinion in neurobiology, vol. 23, no. 2, pages 187–194, 2013. (Cited on page 2.)

[Markov *et al.* 2013] Nikola T Markov, Mária Ercsey-Ravasz, David C Van Essen, Kenneth Knoblauch, Zoltán Toroczkai and Henry Kennedy. *Cortical high-density counterstream architectures.* Science, vol. 342, no. 6158, page 1238406, 2013. (Cited on page 2.)

[Martens & Sutskever 2011] James Martens and Ilya Sutskever. *Learning recurrent neural networks with hessian-free optimization.* In ICML, 2011. (Cited on page 19.)

[Martens *et al.* 2010] James Martens*et al. Deep learning via hessian-free optimization.* In ICML, volume 27, pages 735–742, 2010. (Cited on page 19.)

[McCauley & Christiansen 2019] Stewart M McCauley and Morten H Christiansen. *Language learning as language use: A cross-linguistic model of child language development.* Psychological review, vol. 126, no. 1, page 1, 2019. (Cited on page 58.)

[McGeer *et al.* 1990] Tad McGeer*et al. Passive dynamic walking.* Int. J. Robotics Res., vol. 9, no. 2, pages 62–82, 1990. (Cited on page 75.)

[McInnes *et al.* 2018] Leland McInnes, John Healy and James Melville. *Umap: Uniform manifold approximation and projection for dimension reduction.* arXiv preprint arXiv:1802.03426, 2018. (Cited on pages 29, 36, 50 and 51.)

[Meister *et al.* 2007] Ingo G. Meister, Stephen M. Wilson, Choi Deblieck, Allan D. Wu and Marco Iacoboni. *The Essential Role of Premotor Cortex in Speech Perception.* Current Biology, vol. 17, no. 19, pages 1692–1696, October 2007. (Cited on page 44.)

[Miikkulainen *et al.* 2019] Risto Miikkulainen, Jason Liang, Elliot Meyerson, Aditya Rawal, Daniel Fink, Olivier Francon, Bala Raju, Hormoz Shahrzad, Arshak Navruzyan, Nigel Duffy*et al. Evolving deep neural networks.* In Artificial intelligence in the age of neural networks and brain computing, pages 293–312. Elsevier, 2019. (Cited on page 65.)

[Mikolov *et al.* 2013] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado and Jeff Dean. *Distributed representations of words and phrases and their compositionality.* In Advances in neural information processing systems, pages 3111–3119, 2013. (Cited on pages 39 and 43.)

[Moro 2014] Andrea Moro. *Response to Pulvermüller: the syntax of actions and other metaphors.* Trends in Cognitive Sciences, vol. 18, no. 5, page 221, 2014. (Cited on page 1.)

[Moulin-Frier *et al.* 2014] C Moulin-Frier, SM Nguyen and PY Oudeyer. *Self-organization of early vocal development in infants and machines: the role of intrinsic motivation.* Frontiers in psychology, vol. 4, page 1006, 2014. (Cited on page 56.)

[Moulin-Frier *et al.* 2015] Clément Moulin-Frier, Julien Diard, Jean-Luc Schwartz and Pierre Bessière. *COSMO ("Communicating about Objects using Sensory–Motor Operations"): A Bayesian modeling framework for studying speech*

communication and the emergence of phonological systems. Journal of Phonetics, vol. 53, pages 5–41, 2015. (Cited on pages 45 and 55.)

[Nabé *et al.* 2021] Mamady Nabé, Jean-Luc Schwartz and Julien Diard. *COSMO-Onset: A Neurally-Inspired Computational Model of Spoken Word Recognition, Combining Top-Down Prediction and Bottom-Up Detection of Syllabic Onsets*. Frontiers in Systems Neuroscience, page 75, 2021. (Cited on page 55.)

[Nabé *et al.* 2022] Mamady Nabé, Jean-Luc Schwartz and Julien Diard. *Bayesian gates: a probabilistic modeling tool for temporal segmentation of sensory streams into sequences of perceptual accumulators*. In 44th Annual Conference of the Cognitive Science Society, 2022. (Cited on page 55.)

[Nagel 1974] Thomas Nagel. *What is it like to be a bat?* The philosophical review, vol. 83, no. 4, pages 435–450, 1974. (Cited on pages 69 and 72.)

[Nastase *et al.* 2021] Samuel A. Nastase, Yun-Fei Liu, Hanna Hillman, Asieh Zadbood, Liat Hasenfratz, Neggin Keshavarzian, Janice Chen, Christopher J. Honey, Yaara Yeshurun, Mor Regev, Mai Nguyen, Claire H. C. Chang, Christopher Baldassano, Olga Lositsky, Erez Simony, Michael A. Chow, Yuan Chang Leong, Paula P. Brooks, Emily Micciche, Gina Choe, Ariel Goldstein, Tamara Vanderwal, Yaroslav O. Halchenko, Kenneth A. Norman and Uri Hasson. *The "Narratives" fMRI dataset for evaluating models of naturalistic language comprehension*. Scientific Data, vol. 8, no. 1, September 2021. (Cited on page 54.)

[Nelson *et al.* 2017] Matthew J. Nelson, Imen El Karoui, Kristof Giber, Xiaofang Yang, Laurent Cohen, Hilda Koopman, Sydney S. Cash, Lionel Naccache, John T. Hale, Christophe Pallier and Stanislas Dehaene. *Neurophysiological dynamics of phrase-structure building during sentence processing*. Proceedings of the National Academy of Sciences, vol. 114, no. 18, pages E3669–E3678, 2017. (Cited on page 54.)

[Nemeth *et al.* 2013] Erwin Nemeth, Nadia Pieretti, Sue Anne Zollinger, Nicole Geberzahn, Jesko Partecke, Ana Catarina Miranda and Henrik Brumm. *Bird song and anthropogenic noise: vocal constraints may explain why birds sing higher-frequency songs in cities*. Proceedings of the Royal Society B: Biological Sciences, vol. 280, no. 1754, page 20122798, March 2013. (Cited on page 68.)

[Oota *et al.* 2022] Subba Reddy Oota, Frédéric Alexandre and Xavier Hinaut. Cross-Situational Learning Towards Robot Grounding. HAL preprint, April 2022. (Cited on pages 3, 21, 25, 36, 49, 59 and 63.)

[O'Regan 2011] J Kevin O'Regan. Why red doesn't sound like a bell: Understanding the feel of consciousness. OUP USA, 2011. (Cited on pages ii and 73.)

[Pagliarini *et al.* 2021a] Silvia Pagliarini, Arthur Leblois and Xavier Hinaut. *Canary Vocal Sensorimotor Model with RNN Decoder and Low-dimensional GAN Generator*. In ICDL 2021- IEEE International Conference on Development and Learning, Beijing, China, August 2021. (Cited on pages 3, 51, 53, 55, 56, 57 and 65.)

[Pagliarini *et al.* 2021b] Silvia Pagliarini, Arthur Leblois and Xavier Hinaut. *Vocal Imitation in Sensorimotor Learning Models: A Comparative Review*. IEEE Transactions on Cognitive and Developmental Systems, vol. 13, no. 2, pages 326–342, June 2021. (Cited on pages 3, 44, 55, 56, 57 and 64.)

[Pagliarini *et al.* 2021c] Silvia Pagliarini, Nathan Trouvain, Arthur Leblois and Xavier Hinaut. *What does the Canary Say? Low-Dimensional GAN Applied to Birdsong*. HAL preprint ⟨hal-03244723v2⟩, 2021. (Cited on pages 3, 50, 51, 53, 55, 56, 57 and 65.)

[Pallier *et al.* 2011] C. Pallier, A.-D. Devauchelle and S. Dehaene. *Cortical representation of the constituent structure of sentences*. Proceedings of the National Academy of Sciences, vol. 108, no. 6, pages 2522–2527, January 2011. (Cited on page 54.)

[Pascanu *et al.* 2013] Razvan Pascanu, Tomas Mikolov and Yoshua Bengio. *On the difficulty of training recurrent neural networks*. In International conference on machine learning, pages 1310–1318. PMLR, 2013. (Cited on pages 18 and 19.)

[Pastra & Aloimonos 2012] Katerina Pastra and Yiannis Aloimonos. *The minimalist grammar of action*. Philosophical Transactions of the Royal Society B: Biological Sciences, vol. 367, no. 1585, pages 103–117, 2012. (Cited on page 1.)

[Pattee & Rączaszek-Leonardi 2012] Howard Hunt Pattee and Joanna Rączaszek-Leonardi. Laws, language and life: Howard pattee's classic papers on the physics of symbols with contemporary commentary, volume 7. Springer Science & Business Media, 2012. (Cited on pages 69 and 75.)

[Pedrelli & Hinaut 2020] Luca Pedrelli and Xavier Hinaut. *Hierarchical-Task Reservoir for Anytime POS Tagging from Continuous Speech*. In 2020 International Joint Conference on Neural Networks (IJCNN). IEEE, July 2020. (Cited on pages 3, 22, 46, 47, 48 and 52.)

[Pedrelli & Hinaut 2022] Luca Pedrelli and Xavier Hinaut. *Hierarchical-Task Reservoir for Online Semantic Analysis From Continuous Speech*. IEEE Transactions on Neural Networks and Learning Systems, vol. 33, no. 6, pages 2654–2663, June 2022. (Cited on pages 3, 27, 46, 47 and 52.)

[Pfeifer & Bongard 2006] Rolf Pfeifer and Josh Bongard. How the body shapes the way we think: a new view of intelligence. MIT press, 2006. (Cited on page 70.)

[Pfeifer & Pitti 2012] Rolf Pfeifer and Alexandre Pitti. La révolution de l'intelligence du corps. Manuella éd., 2012. (Cited on pages ii, 70 and 75.)

[Philippsen *et al.* 2014] AK Philippsen, RF Reinhart and B Wrede. *Learning how to speak: Imitation-based refinement of syllable production in an articulatory-acoustic model.* In ICDL-EpiRob, pages 195–200. IEEE, 2014. (Cited on page 55.)

[Philippsen *et al.* 2016] Anja Kristina Philippsen, René Felix Reinhart and Britta Wrede. *Goal babbling of acoustic-articulatory models with adaptive exploration noise.* In 2016 Joint IEEE International Conference on Development and Learning and Epigenetic Robotics (ICDL-EpiRob), pages 72–78. IEEE, 2016. (Cited on page 55.)

[Philippsen 2021] Anja Philippsen. *Goal-directed exploration for learning vowels and syllables: a computational model of speech acquisition.* KI-Künstliche Intelligenz, vol. 35, no. 1, pages 53–70, 2021. (Cited on pages 55 and 56.)

[Pickering & Garrod 2013] Martin J. Pickering and Simon Garrod. *An integrated theory of language production and comprehension.* Behavioral and Brain Sciences, vol. 36, no. 4, pages 329–347, June 2013. (Cited on page 39.)

[Pitti *et al.* 2020] Alexandre Pitti, Mathias Quoy, Catherine Lavandier and Sofiane Boucenna. *Gated spiking neural network using Iterative Free-Energy Optimization and rank-order coding for structure learning in memory sequences (INFERNO GATE).* Neural Networks, vol. 121, pages 242–258, 2020. (Cited on pages 43 and 58.)

[Pitti *et al.* 2022] Alexandre Pitti, Claudio Weidmann and Mathias Quoy. *Digital computing through randomness and order in neural networks.* Proceedings of the National Academy of Sciences, vol. 119, no. 33, page e2115335119, 2022. (Cited on page 58.)

[Prather *et al.* 2008] Jonathan F Prather, Susan Peters, Stephen Nowicki and Richard Mooney. *Precise auditory–vocal mirroring in neurons for learned vocal communication.* Nature, vol. 451, no. 7176, pages 305–310, 2008. (Cited on page 2.)

[Pulvermüller & Fadiga 2010] Friedemann Pulvermüller and Luciano Fadiga. *Active perception: sensorimotor circuits as a cortical basis for language.* Nature reviews neuroscience, vol. 11, no. 5, pages 351–360, 2010. (Cited on pages 39 and 44.)

[Pulvermüller 2013] Friedemann Pulvermüller. *How neurons make meaning: brain mechanisms for embodied and abstract-symbolic semantics.* Trends in Cognitive Sciences, vol. 17, no. 9, pages 458–470, September 2013. (Cited on page 40.)

[Pulvermüller 2014] Friedemann Pulvermüller. *The syntax of action.* Trends in Cognitive Sciences, vol. 18, no. 5, pages 219–220, May 2014. (Cited on page 1.)

[Rączaszek-Leonardi & Kelso 2008] Joanna Rączaszek-Leonardi and JA Scott Kelso. *Reconciling symbolic and dynamic aspects of language: Toward a dynamic psycholinguistics.* New ideas in psychology, vol. 26, no. 2, pages 193–207, 2008. (Cited on page 69.)

[Rączaszek-Leonardi *et al.* 2018] Joanna Rączaszek-Leonardi, Iris Nomikou, Katharina J Rohlfing and Terrence W Deacon. *Language development from an ecological perspective: Ecologically valid ways to abstract symbols.* Ecological Psychology, vol. 30, no. 1, pages 39–73, 2018. (Cited on page 68.)

[Radford *et al.* 2019] Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, Ilya Sutskever*et al. Language models are unsupervised multitask learners.* OpenAI blog, vol. 1, no. 8, page 9, 2019. (Cited on page 43.)

[Rae *et al.* 2016] Jack Rae, Jonathan J Hunt, Ivo Danihelka, Timothy Harley, Andrew W Senior, Gregory Wayne, Alex Graves and Timothy Lillicrap. *Scaling memory-augmented neural networks with sparse reads and writes.* Advances in Neural Information Processing Systems, vol. 29, 2016. (Cited on page 63.)

[Raffel *et al.* 2020] Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, Peter J Liu*et al. Exploring the limits of transfer learning with a unified text-to-text transformer.* J. Mach. Learn. Res., vol. 21, no. 140, pages 1–67, 2020. (Cited on page 43.)

[Regier & Kay 2009] Terry Regier and Paul Kay. *Language, thought, and color: Whorf was half right.* Trends in cognitive sciences, vol. 13, no. 10, pages 439–446, 2009. (Cited on page 74.)

[Rigotti *et al.* 2013] Mattia Rigotti, Omri Barak, Melissa R Warden, Xiao-Jing Wang, Nathaniel D Daw, Earl K Miller and Stefano Fusi. *The importance of mixed selectivity in complex cognitive tasks.* Nature, vol. 497, no. 7451, pages 585–590, 2013. (Cited on page 19.)

[Rizzolatti & Arbib 1998] Giacomo Rizzolatti and Michael A. Arbib. *Language within our grasp.* Trends in Neurosciences, vol. 21, no. 5, pages 188–194, May 1998. (Cited on page 2.)

[Rodan & Tino 2010] Ali Rodan and Peter Tino. *Minimum complexity echo state network.* IEEE transactions on neural networks, vol. 22, no. 1, pages 131–144, 2010. (Cited on page 18.)

[Rolf *et al.* 2010] Matthias Rolf, Jochen J Steil and Michael Gienger. *Goal babbling permits direct learning of inverse kinematics.* IEEE Transactions on Autonomous Mental Development, vol. 2, no. 3, pages 216–229, 2010. (Cited on page 56.)

[Rougier & Boniface 2011] Nicolas Rougier and Yann Boniface. *Dynamic self-organising map.* Neurocomputing, vol. 74, no. 11, pages 1840–1847, 2011. (Cited on page 58.)

[Rougier *et al.* 2005] Nicolas P Rougier, David C Noelle, Todd S Braver, Jonathan D Cohen and Randall C O'Reilly. *Prefrontal cortex and flexible cognitive control: Rules without symbols.* Proceedings of the National Academy of Sciences, vol. 102, no. 20, pages 7338–7343, 2005. (Cited on page 54.)

[Roy 2002] Deb K. Roy. *Learning visually grounded words and syntax for a scene description task.* Computer Speech & Language, vol. 16, no. 3-4, pages 353–385, July 2002. (Cited on page 45.)

[Schlenker 2017] Philippe Schlenker. *Sign Language and the Foundations of Anaphora.* Annual Review of Linguistics, vol. 3, no. 1, pages 149–177, January 2017. (Cited on page 1.)

[Schrauwen *et al.*] Benjamin Schrauwen, Marion Wardermann, David Verstraeten, Jochen J. Steil and Dirk Stroobandt. *Improving Reservoirs Using Intrinsic Plasticity.* vol. 71, no. 7, pages 1159–1171. (Cited on page 43.)

[Schwartz *et al.* 2012] Jean-Luc Schwartz, Anahita Basirat, Lucie Ménard and Marc Sato. *The Perception-for-Action-Control Theory (PACT): A perceptuo-motor theory of speech perception.* Journal of Neurolinguistics, vol. 25, no. 5, pages 336–354, September 2012. (Cited on pages 39, 44, 45 and 55.)

[Sejnowski 2022] Terrence Sejnowski. *Large Language Models and the Reverse Turing Test.* arXiv preprint arXiv:2207.14382, 2022. (Cited on pages 74 and 75.)

[Steels 1998] Luc Steels. *The origins of syntax in visually grounded robotic agents.* Artificial Intelligence, vol. 103, no. 1-2, pages 133–156, August 1998. (Cited on page 45.)

[Steil 2007] Jochen J Steil. *Online reservoir adaptation by intrinsic plasticity for backpropagation–decorrelation and echo state learning.* Neural networks, vol. 20, no. 3, pages 353–364, 2007. (Cited on page 43.)

[Sterzer *et al.* 2009] Philipp Sterzer, Andreas Kleinschmidt and Geraint Rees. *The neural bases of multistable perception.* Trends in cognitive sciences, vol. 13, no. 7, pages 310–318, 2009. (Cited on page 41.)

[Stramandinoli *et al.* 2012] Francesca Stramandinoli, Davide Marocco and Angelo Cangelosi. *The grounding of higher order concepts in action and language: A cognitive robotics model.* Neural Networks, vol. 32, pages 165–173, August 2012. (Cited on page 45.)

[Strock *et al.* 2020] Anthony Strock, Xavier Hinaut and Nicolas P. Rougier. *A Robust Model of Gated Working Memory.* Neural Computation, vol. 32, no. 1, pages 153–181, January 2020. (Cited on pages 3, 53, 63, 64 and 69.)

[Strock *et al.* 2022] Anthony Strock, Nicolas P. Rougier and Xavier Hinaut. *Latent Space Exploration and Functionalization of a Gated Working Memory Model Using Conceptors.* Cognitive Computation, January 2022. (Cited on pages 3 and 64.)

[Sugita & Tani 2005] Yuuya Sugita and Jun Tani. *Learning Semantic Combinatoriality from the Interaction between Linguistic and Behavioral Processes.* Adaptive Behavior, vol. 13, no. 1, pages 33–52, March 2005. (Cited on page 45.)

[Sun & Alexandre 2013] Ron Sun and Frederic Alexandre. Connectionist-symbolic integration: From unified to hybrid approaches. Psychology Press, 2013. (Cited on page 42.)

[Sussillo & Abbott 2009] David Sussillo and L.F. Abbott. *Generating Coherent Patterns of Activity from Chaotic Neural Networks.* Neuron, vol. 63, no. 4, pages 544–557, August 2009. (Cited on pages 47 and 49.)

[Sutskever *et al.* 2011] Ilya Sutskever, James Martens and Geoffrey E Hinton. *Generating text with recurrent neural networks.* In ICML, 2011. (Cited on page 19.)

[Tanaka *et al.* 2019] Gouhei Tanaka, Toshiyuki Yamane, Jean Benoit Héroux, Ryosho Nakane, Naoki Kanazawa, Seiji Takeda, Hidetoshi Numata, Daiju Nakano and Akira Hirose. *Recent advances in physical reservoir computing: A review.* Neural Networks, vol. 115, pages 100–123, 2019. (Cited on page 20.)

[Taniguchi *et al.* 2016] Tadahiro Taniguchi, Takayuki Nagai, Tomoaki Nakamura, Naoto Iwahashi, Tetsuya Ogata and Hideki Asoh. *Symbol emergence in robotics: a survey.* Advanced Robotics, vol. 30, no. 11-12, pages 706–728, April 2016. (Cited on pages 40, 45 and 68.)

[Taniguchi *et al.* 2017] Akira Taniguchi, Tadahiro Taniguchi and Angelo Cangelosi. *Cross-situational learning with Bayesian generative models for multimodal category and word learning in robots.* Frontiers in neurorobotics, vol. 11, page 66, 2017. (Cited on page 45.)

[Taniguchi *et al.* 2018] Tadahiro Taniguchi, Emre Ugur, Matej Hoffmann, Lorenzo Jamone, Takayuki Nagai, Benjamin Rosman, Toshihiko Matsuka, Naoto Iwahashi, Erhan Oztop, Justus Piater*et al. Symbol emergence in cognitive developmental systems: a survey.* IEEE transactions on Cognitive and Developmental Systems, vol. 11, no. 4, pages 494–516, 2018. (Cited on page 68.)

[Taniguchi *et al.* 2022] Tadahiro Taniguchi, Yuto Yoshida, Akira Taniguchi and Yoshinobu Hagiwara. *Emergent Communication through Metropolis-Hastings Naming Game with Deep Generative Models.* arXiv preprint arXiv:2205.12392, 2022. (Cited on page 68.)

[Thoenissen *et al.* 2002] Daniel Thoenissen, Karl Zilles and Ivan Toni. *Differential Involvement of Parietal and Precentral Regions in Movement Preparation and Motor Intention.* The Journal of Neuroscience, vol. 22, no. 20, pages 9024–9034, October 2002. (Cited on page 1.)

[Thoppilan *et al.* 2022] Romal Thoppilan, Daniel De Freitas, Jamie Hall, Noam Shazeer, Apoorv Kulshreshtha, Heng-Tze Cheng, Alicia Jin, Taylor Bos, Leslie Baker, Yu Du*et al. Lamda: Language models for dialog applications.* arXiv preprint arXiv:2201.08239, 2022. (Cited on page 73.)

[Tomasello 2003] Michael Tomasello. Constructing a language: A usage-based theory of language acquisition. Harvard University Press, 2003. (Cited on page 56.)

[Tourville & Guenther 2011] JA Tourville and FH Guenther. *The DIVA model: A neural theory of speech acquisition and production.* Language and cognitive processes, vol. 26, no. 7, pages 952–981, 2011. (Cited on pages 44 and 56.)

[Triefenbach *et al.* 2013] Fabian Triefenbach, Azarakhsh Jalalvand, Kris Demuynck and Jean-Pierre Martens. *Acoustic modeling with hierarchical reservoirs.* IEEE Transactions on Audio, Speech, and Language Processing, vol. 21, no. 11, pages 2439–2450, 2013. (Cited on page 47.)

[Trouvain & Hinaut 2021] Nathan Trouvain and Xavier Hinaut. *Canary Song Decoder: Transduction and Implicit Segmentation with ESNs and LTSMs.* In ICANN 2021 - 30th International Conference on Artificial Neural Networks, volume 12895 of *Farkaš I., Masulli P., Otte S., Wermter S. (eds) Artificial Neural Networks and Machine Learning – ICANN 2021. Lecture Notes in Computer Science*, pages 71–82, Bratislava, Slovakia, September 2021. Springer, Cham. (Cited on pages 3, 21, 52, 57 and 65.)

[Trouvain & Hinaut 2022] Nathan Trouvain and Xavier Hinaut. reservoirpy: A Simple and Flexible Reservoir Computing Tool in Python. hal-03699931 preprint, June 2022. (Cited on pages 3, 33 and 58.)

[Trouvain *et al.* 2020] Nathan Trouvain, Luca Pedrelli, Thanh Trung Dinh and Xavier Hinaut. *ReservoirPy: an Efficient and User-Friendly Library to Design Echo State Networks.* In ICANN 2020 - 29th International Conference on Artificial Neural Networks, Bratislava, Slovakia, September 2020. (Cited on pages 3, 33, 58 and 60.)

[Trouvain *et al.* 2022] Nathan Trouvain, Nicolas P. Rougier and Xavier Hinaut. *Create Efficient and Complex Reservoir Computing Architectures with ReservoirPy.* In SAB 2022 - FROM ANIMALS TO ANIMATS 16: The 16th International Conference on the Simulation of Adaptive Behavior, Cergy-Pontoise / Hybrid, France, September 2022. (Cited on pages 3, 34 and 58.)

[Van Hell *et al.* 2015] Janet G Van Hell, Kaitlyn A Litcofsky and Caitlin Y Ting. *Intra-sentential code-switching: Cognitive and neural approaches.* The Cambridge handbook of bilingual processing, pages 459–482, 2015. (Cited on page 54.)

[VanDam *et al.* 2016] Mark VanDam, Anne S Warlaumont, Elika Bergelson, Alejandrina Cristia, Melanie Soderstrom, Paul De Palma and Brian MacWhinney. *HomeBank: An online repository of daylong child-centered audio recordings.* In Seminars in speech and language, volume 37, pages 128–142. Thieme Medical Publishers, 2016. (Cited on page 55.)

[Vapnik 1999] Vladimir Vapnik. The nature of statistical learning theory. Springer science & business media, 1999. (Cited on page 20.)

[Variengien & Hinaut 2020] Alexandre Variengien and Xavier Hinaut. *A journey in ESN and LSTM visualisations on a language task.* arXiv, page arXiv:2012.01748, 2020. (Cited on pages 3, 21, 25, 35, 47, 49, 51, 53, 54, 59 and 63.)

[Vaswani *et al.* 2017] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Ł ukasz Kaiser and Illia Polosukhin. *Attention is All you Need.* In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan and R. Garnett, editors, Advances in Neural Information Processing Systems, volume 30. Curran Associates, Inc., 2017. (Cited on pages 43 and 63.)

[Verstraeten 2009] David Verstraeten. *Reservoir Computing: computation with dynamical systems.* PhD thesis, Ghent University, 2009. (Cited on pages 20 and 23.)

[Voegtlin 2002] Thomas Voegtlin. *Recursive self-organizing maps.* Neural networks, vol. 15, no. 8-9, pages 979–991, 2002. (Cited on page 65.)

[Warlaumont & Finnegan 2016] Anne S Warlaumont and Megan K Finnegan. *Learning to produce syllabic speech sounds via reward-modulated neural plasticity.* PloS one, vol. 11, no. 1, page e0145096, 2016. (Cited on page 57.)

[Warlaumont *et al.* 2013] Anne S Warlaumont, Gert Westermann, Eugene H Buder and D Kimbrough Oller. *Prespeech motor learning in a neural network using reinforcement.* Neural Networks, vol. 38, pages 64–75, 2013. (Cited on page 57.)

[Werbos 1988] Paul J Werbos. *Generalization of backpropagation with application to a recurrent gas market model.* Neural networks, vol. 1, no. 4, pages 339–356, 1988. (Cited on page 18.)

[Werbos 1990] Paul J Werbos. *Backpropagation through time: what it does and how to do it.* Proceedings of the IEEE, vol. 78, no. 10, pages 1550–1560, 1990. (Cited on page 18.)

[Yamada *et al.* 2016] Tatsuro Yamada, Shingo Murata, Hiroaki Arie and Tetsuya Ogata. *Dynamical integration of language and behavior in a recurrent neural network for human–robot interaction.* Frontiers in neurorobotics, vol. 10, page 5, 2016. (Cited on page 61.)

[Yamashita & Tani 2008] Yuichi Yamashita and Jun Tani. *Emergence of Functional Hierarchy in a Multiple Timescale Neural Network Model: A Humanoid Robot Experiment.* PLoS Computational Biology, vol. 4, no. 11, page e1000220, November 2008. (Cited on page 45.)

[Zacks *et al.* 2007] Jeffrey M. Zacks, Nicole K. Speer, Khena M. Swallow, Todd S. Braver and Jeremy R. Reynolds. *Event perception: A mind-brain perspective.* Psychological Bulletin, vol. 133, no. 2, pages 273–293, March 2007. (Cited on page 63.)

[Zajzon *et al.* 2018] Barna Zajzon, Renato Duarte and Abigail Morrison. *Transferring state representations in hierarchical spiking neural networks.* In 2018 International Joint Conference on Neural Networks (IJCNN), pages 1–9. IEEE, 2018. (Cited on page 64.)

[Zajzon *et al.* 2022] Barna Zajzon, David Dahmen, Abigail Morrison and Renato Duarte. *Signal denoising through topographic modularity of neural circuits.* bioRxiv, 2022. (Cited on page 64.)

[Zatorre *et al.* 2007] Robert J. Zatorre, Joyce L. Chen and Virginia B. Penhune. *When the brain plays music: auditory–motor interactions in music perception and production.* Nature Reviews Neuroscience, vol. 8, no. 7, pages 547–558, July 2007. (Cited on page 2.)