A Vocal Learning Model: Development And Analysis

Silvia Pagliarini, Arthur Leblois, Xavier Hinaut

A Vocal Learning Model: Development And Analysis

Silvia Pagliarini[1,2,3], Arthur Leblois[3], Xavier Hinaut[1,2,3]

1) Mnemosyne, Inria Bordeaux Sud-Ouest, UMR 5800, CNR, Bordeaux INP, Talence, France
2) LaBRI, UMR 5800, CNRS, University of Bordeaux, Talence, France
3) IMN, UMR 5293, CNRS, University of Bordeaux, Bordeaux, France

AIM: build a vocal learning model underlying song learning in birds, and understand how to make it biologically plausible.

INTRODUCTION

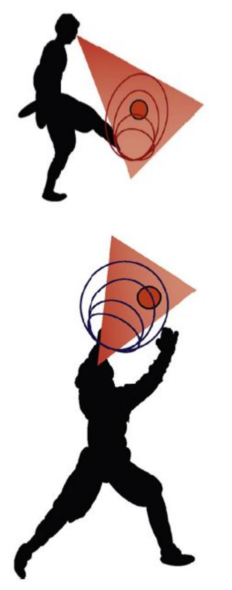
KEYWORDS

Sensorimotor learning: control problem which maps a sensory input into a motor output.

Imitation: the ability to replicate another's behaviour, i.e. to generate an appropriate motor command to reproduce a sensory stimulus.

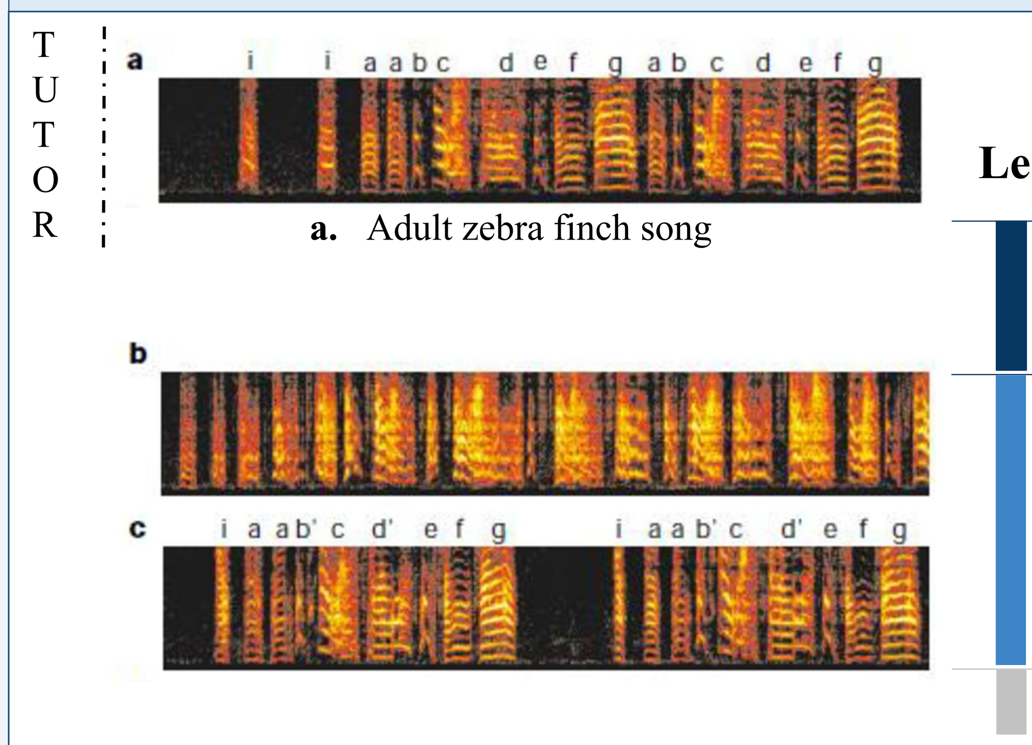
Inverse model learning: use the desired and actual motor configuration to estimate the motor commands needed to reach the desired configuration (to produce the appropriate sensory stimulus).

Reinforcement learning: learn an action policy to maximize the expected reward (which encode the goal of the learning).



Da Cunha et al.,2010

VOCAL LEARNING IN BIRDS

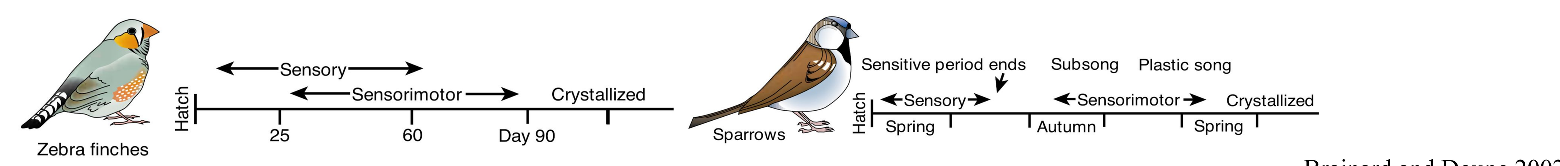


Learning phases:

Sensory phase
Building neural representation of the song.

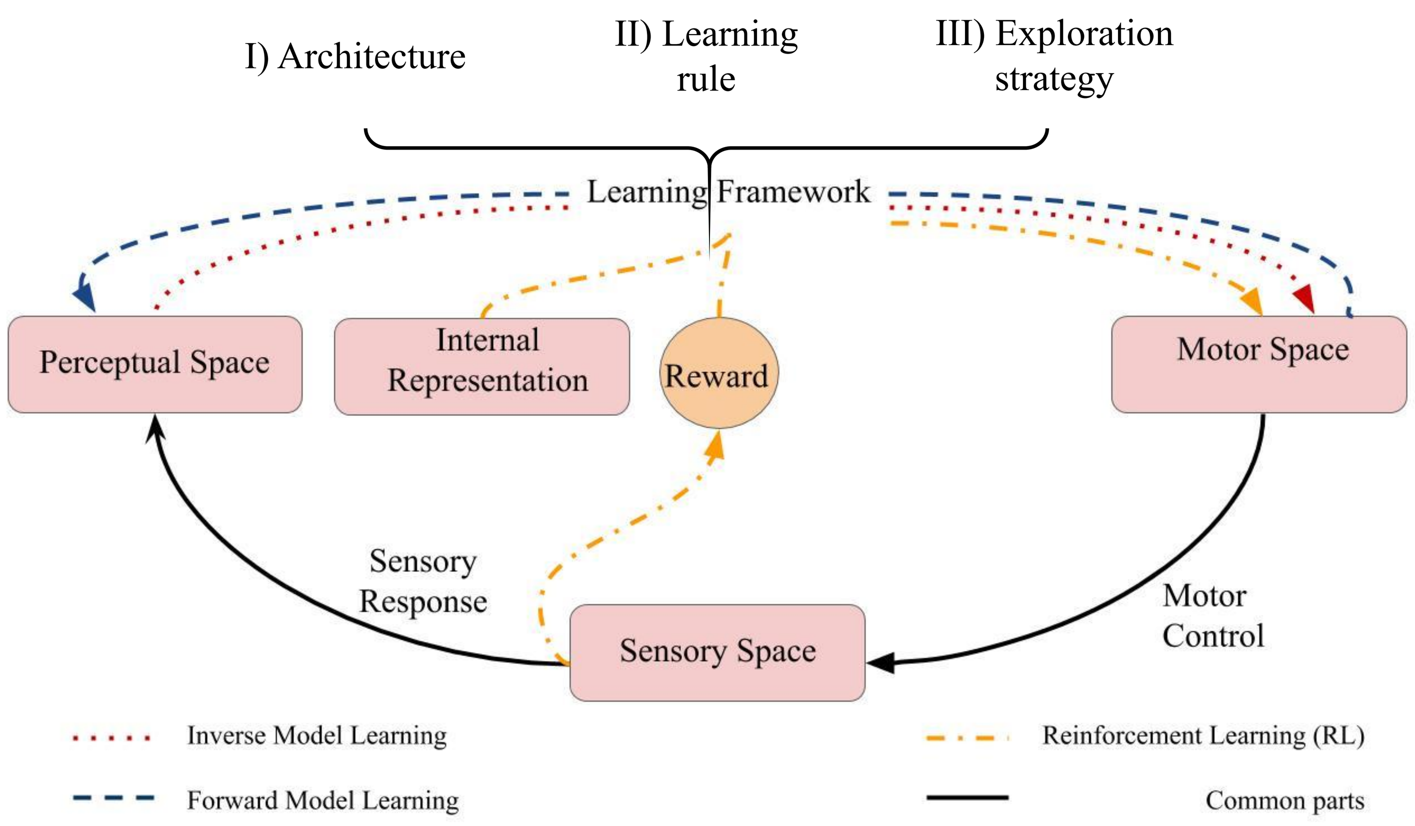
Sensorimotor phase
b. Subsong (babbling).

c. Plastic song (temporally structured).
Crystallization



Brainard and Doupe,2002

SENSORIMOTOR MODEL COMPONENTS [1]

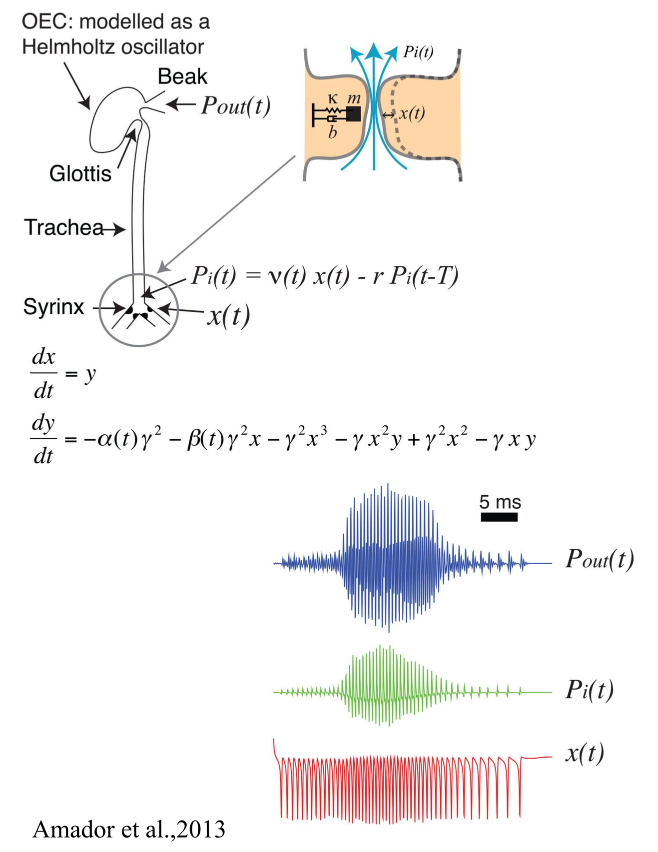


Motor control model

- Respiratory system and vocal organs.
- Anatomical properties and small size of birds make the investigation of vocal fold mechanisms difficult.
- The vocal output is driven by a complex gesture-dependent control scheme, and the brain does not control each motor control parameter independently [2].

Perceptual space

- How the brain encodes sensory stimulus.
- Highly nonlinear dynamics.
- Low dimensional representation of the sensory space.



Amador et al.,2013

MOTOR CONTROL MODEL

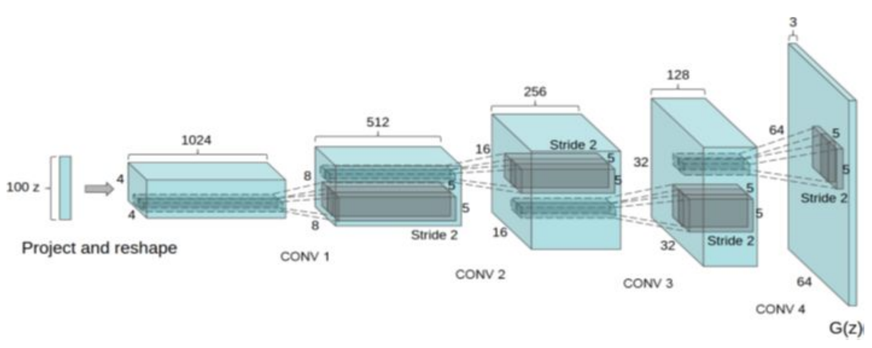
AIM: learn the semantic modes in high-dimensional audio signals, and build a latent space useful for exploration. In addition, have sound production in the model.

WaveGAN [3]

- Two player minimax game: generator VS discriminator [4].
- Inspired by DCGAN architecture [5].

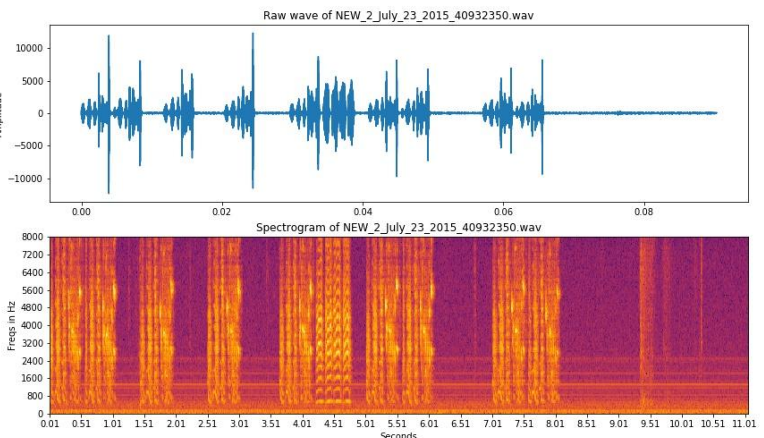
GENERATOR

Random input



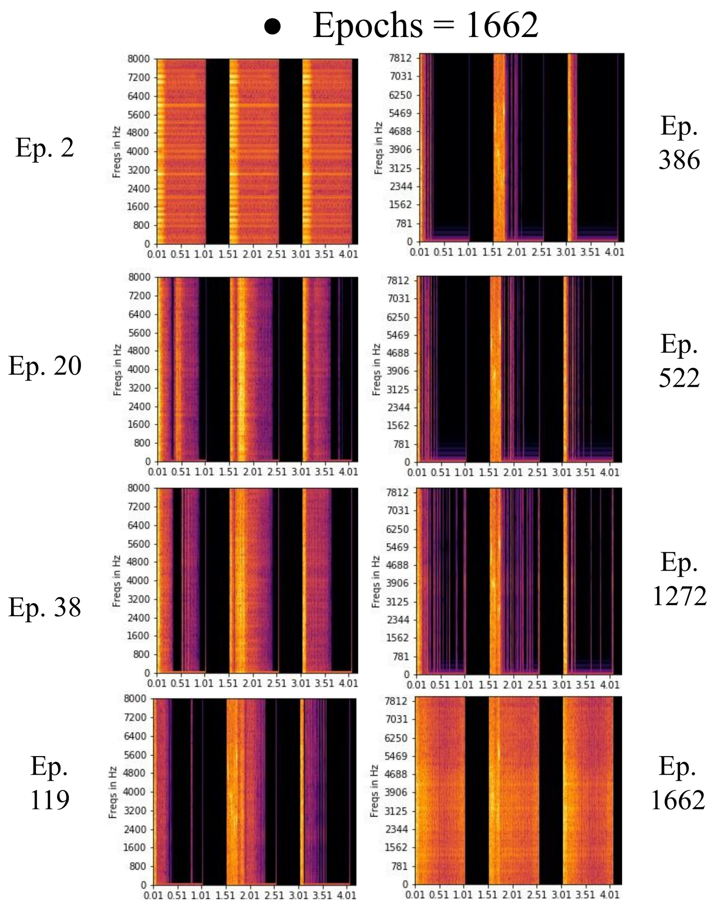
DATASET

- Recordings from an adult zebra finch with sampling rate 44100.
- Downsampled single syllables. N = 4946 syllables.



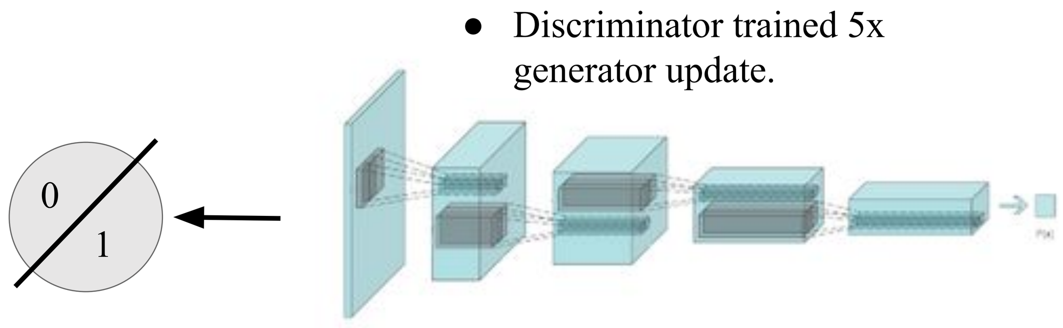
GENERATED SOUND DURING TRAINING

- Batch size = 64
- Epochs = 1662



DISCRIMINATOR

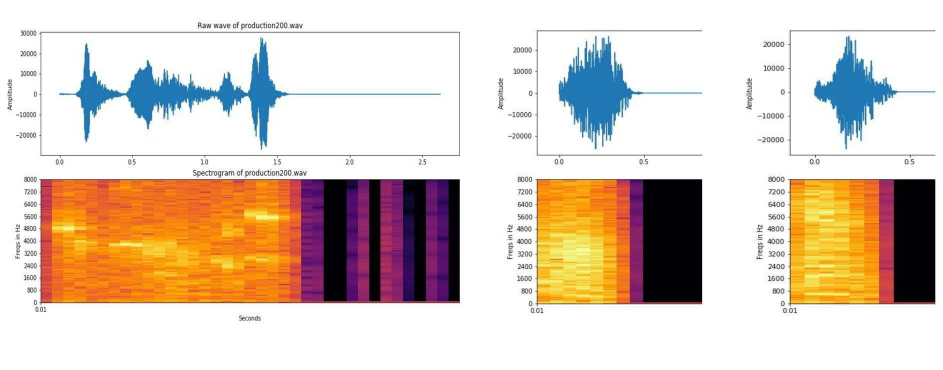
- Discriminator trained 5x generator update.



GENERATED SOUND AFTER TRAINING

- Batch size = 64
- Epochs = 517

Inception Score (IS) = 1.95 ±0.02



PERCEPTUAL SPACE

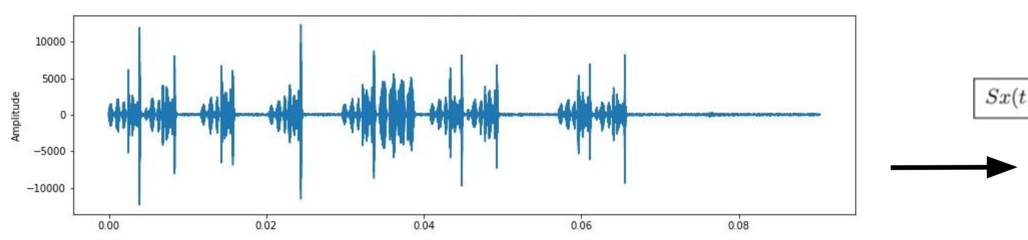
AIM: have a lower dimensional representation of the physical acoustic space, which is the sensory space (sound).

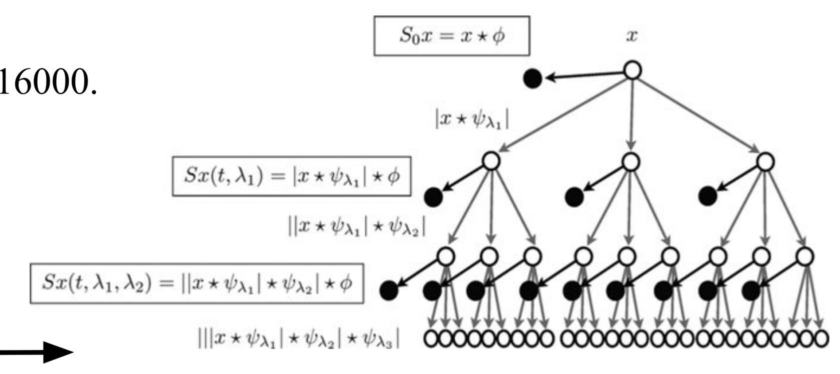
Deep scattering transform [6]

- From MFCC coefficients introducing wavelets.
- Scattering coefficients can be seen as a convolutional neural networks.

DATASET

- Recordings from an adult zebra finch with sampling rate 16000.

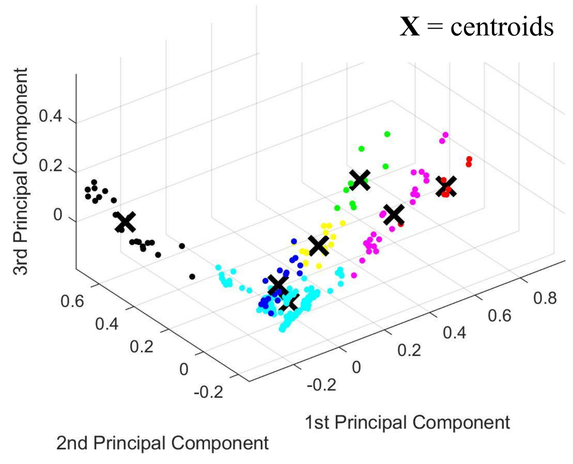




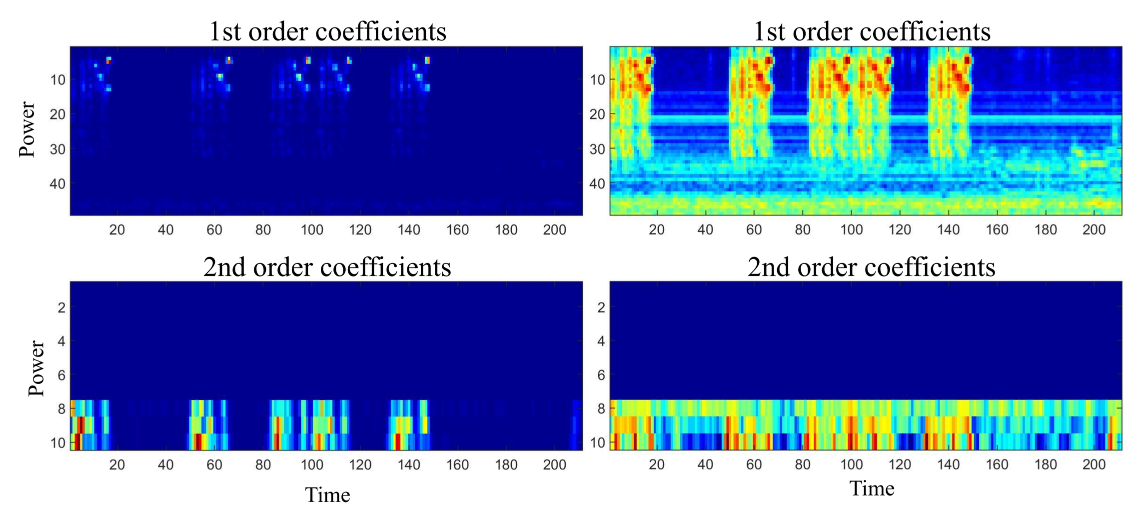
- Averaging window = 800
- 1st and 2nd order coefficients.
- Cell array: elements correspond to layers in the scattering transform.

PCA analysis

- N = 52 syllables.
- PCA analysis.
- k-means algorithm.



Second-order, translation-invariant scattering transform of one-dimensional signals



PERSPECTIVES

Motor control model

- Explore the possibility to use a dynamical system model.
- More exaustive evaluation of the generative modle.

Exploration strategy

- Intrinsic motivation to drive motor exploration.
- Investigate the geometrical properties of latent space generated with the GAN.

BIBLIOGRAPHY

[1] Oudeyer, P. Y. (2005). The self-organization of speech sounds. Journal of Theoretical Biology, 233(3), 435-449.
[2] Elemans, C. P. H., et al. Universal mechanisms of sound production and control in birds and mammals. Nature communications, 2015, 6: 8978.
[3] Donahue, C., McAuley, J., & Puckette, M. (2019). Adversarial Audio Synthesis. ICLR.
[4] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2014). Generative adversarial nets. In Advances in neural information processing systems (pp. 2672-2680).
[5] Radford, A., Metz, L., & Chintala, S. (2015). Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv preprint arXiv:1511.06434.
[6] Andén, J., & Mallat, S. (2014). Deep scattering spectrum. IEEE Transactions on Signal Processing, 62(16), 4114-4128.

AKNOWLEDGMENT: this work was supported by the Inria CORDI-S PhD fellowship grant.