

## Problem Statement

The Zebra Finches are songbirds which learn the song of their tutor. They learn it from 20 days post hatch (DPH) to 80 DPH [1]. Their learning can be split in three phases: i) they have a babbling phase from 20 DPH to 40 DPH, then ii) a plastic song in which protosyllables from the tutor song are produced then iii) a crystallized song at 80 DPH with a fixed song which is a copy of the tutor's song. Zebra finches are commonly used as a model of speech acquisition.

Derégnaucourt, Mitra, Fehér, *et al.* [2] showed that sleep plays an important role in the learning of tutor songs. Indeed, they showed that sleeping has a negative impact on song restitution by zebra finches in the short term but a positive impact on the long run. Indeed, song restitution is less complex and less similar to the tutor song from one morning to the previous day evening, but the greater this loss in performance was overall for one bird, the better this bird will reproduce the tutor song at the end of its learning. Figure 1 shows the negative impact on song similarity the night has.

Dave [3] has found replay sequences of neurones in the motor cortex which correspond to their activity pattern when the birds sing in adult zebra finches during their sleep. This shows that neurones that are highly correlated with bird's own song (BOS) are activated during the night.

Our hypothesis is that during its sleep, the zebra finch restructures the knowledge it has acquired so far. We hypothesize that this restructuring can account for the loss of performance in the short term and an improvement of performance in the long term.

The goal of this internship is to propose a model of the zebra finch song learning which can explain different behavioral data observed such as the correlation between the loss of performance every night and the overall performance at the end of learning, and the different phases of bird song learning (babbling, protosyllables, crystallized song).

## Investigation/Research

We want to build a model which is the most plausible in a real world environment and which is biological plausible. To do so, we will use a bird song synthesizer made by Boari, Perl, Amador, *et al.* [4]. This synthesizer is a biophysical model of zebra finch vocal apparatus.

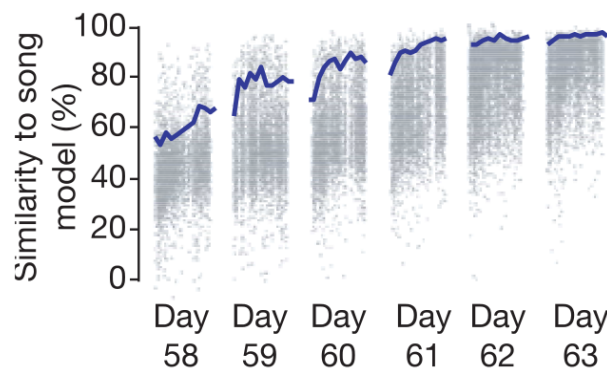


Figure 1: Song similarity from Derégnaucourt, Mitra, Fehér, *et al.* [2]

This synthesizer has the advantage that it be parametrized with relatively few parameters to produce realistic bird songs. As it models the zebra finch vocal apparatus, it is likely that the parameters we send to this synthesizer must be similar to the instructions sent by the zebra finch motor cortex to the vocal apparatus muscles. Indeed, the parameters for this synthesizer is the the labia tension  $\alpha(t)$  and the air-sac pressure  $\beta(t)$  in the apparatus. Boari, Perl, Amador, *et al.* [4] and Amador, Perl, Mindlin, *et al.* [5] have managed to reproduce zebra finches song using this synthesizer. The synthetic songs they produced activated neurons in HVC which are highly selective to bird's own song (BOS). This shows that the synthesized songs are accurate reproduction of BOS.

check  
alpha  
and  
beta  
that I  
keep  
mixing  
up

Boari, Perl, Amador, *et al.* [4] and Amador and Mindlin [6] have already found interesting data by studying the parameters. Indeed, they have found that what can be seen as syllable in the sensory space can be seen as one or several gestures in the parameters space. Indeed, syllable are yield with continuous modification of the parameters. A new syllable will trigger a huge discontinuity in the parameter values.

Amador, Perl, Mindlin, *et al.* [5] even claim to have found correlation with HVC neuron spikes in zebra finches and the gesture trajectory extrema (beginning, maximum, minimum and end) of these gestures. Though, very recent literature shows that these correlates might have been untrue [7], [8]. Even though there is no neural correlate with gesture trajectory extrema (GTE), we hypothesize that GTE identification may play an important role in song learning, as they signal changes in the progression of the parameters through time.

## Proposed Solution

Our goal is to design a simple hill climbing model (see Fig. 2) that fits one specific gesture and a gesture identification algorithm. The gesture identification algorithm will try to segment the tutor song in efficient gestures based on the bird current knowledge. This two step algorithm is similar in some points to an Expectation-Maximization algorithm [9].

So as to do a first gesture segmentation, we will use motor babbling so that the bird robot can explore its sensorimotor space and segment the tutor song based on its babbling. We will then use a interest model to choose which syllable to train [10]. The overview of the whole algorithm is shown in Figure 3. Our idea is that the maximization of the identified gestures occurs during the day and explain the overall advancement in performance.

This part will only cover the learning of gesture, but not the linkage between these gestures. We have yet to find how to learn the pattern of the syllables and the song. This algorithm should be able to use in a smart way the knowledge built by the gesture learning system. The algorithm that learn the syllable transition should also be able to reproduce the different learning strategies that a zebra finch can have. Indeed, the zebra finch can either have a serial strategy, where it only learn one specific syllable at a time, or have a motif strategy, where it learn to reproduce the whole tutor song at every try.

## Expected Implementation

The learning algorithm that we want to suggest must be biologically plausible, therefore we will use the Song Synthesizer from Boari, Perl, Amador, *et al.* [4] to generate real sound-waves. We have already implemented the Python binder to the compiled synthesizer. These sound-waves will then be processed by an auditory system. We plan at first to use Mel-Frequency Cepstrum Coefficients (MFCC), which are used in speech recognition for Humans. Chou, Liu, and Cai [11] showed that MFCC can be used to classify birdsongs. MFCC sums up audio signal in a

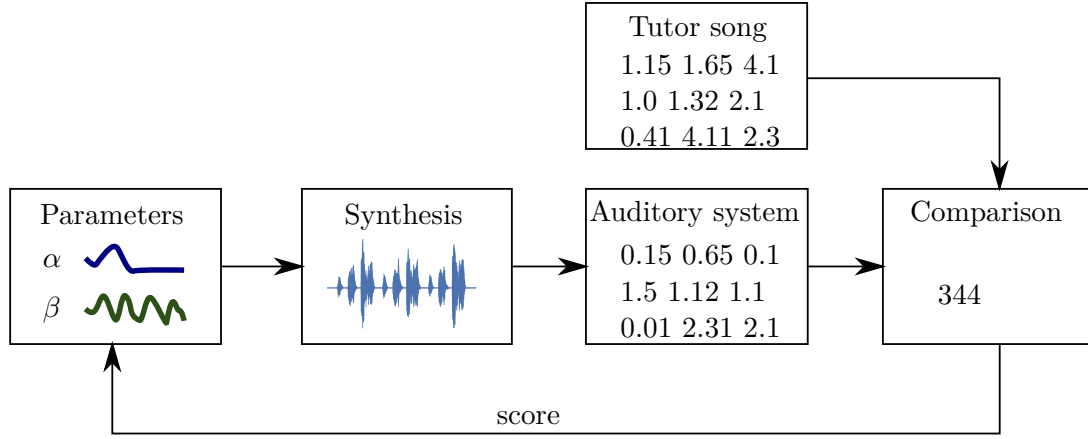


Figure 2: Simple hill climbing model

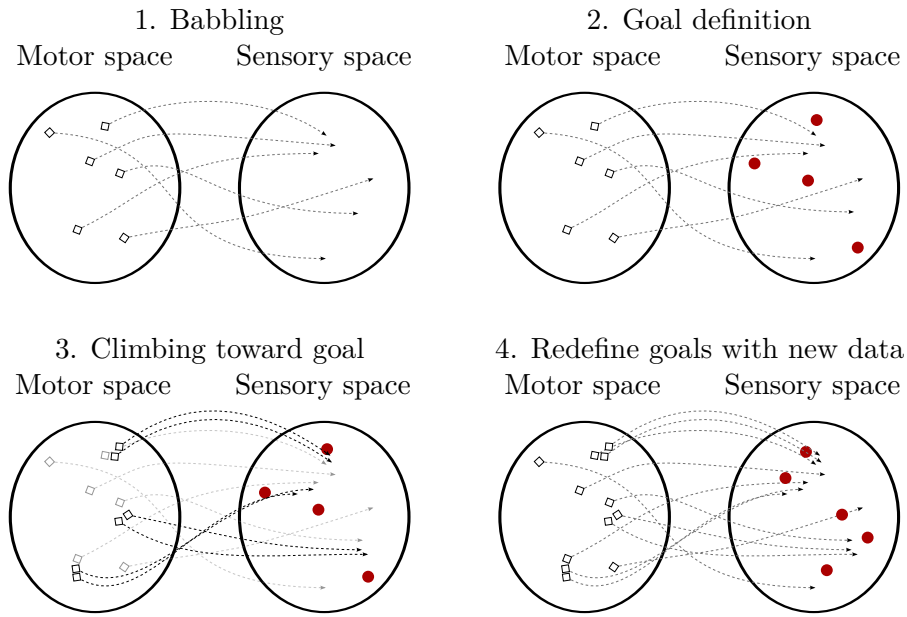


Figure 3: Define goals based on knowledge

few coefficients per time window. Though, we obtained strange results using MFCC squared error between the real song and our generated bird song. Indeed, even though our generated bird songs were qualitatively very bad as they sound like white noise, the difference between their MFCC differences with the real song were lower than the generated song by Boari, Perl, Amador, *et al.* [4] which sound qualitatively very similar. We are currently trying to fine tune the use we have of MFCC and get a better understanding of their behavior. We may change for another sound representation. As the exact length of a syllable may vary from try to try, we plan to use Dynamic Time Warping (DTW) algorithm to avoid error accumulation for a slightly off-time syllable. MFCC and DTW used together are powerful tools for speech recognition [12].

We plan to use a Nearest-Neighbor algorithm to hill climb toward the goals the algorithm defines and remember each try it makes so as to define more realizable goals. The Nearest-Neighbor algorithm and the interest model are available through the Explauto python module [13].

## Analysis & Testing

To assess the quality of our model, we have selected several criteria to meet. First, we want our algorithm to reproduce the results from Derégnaucourt, Mitra, Fehér, *et al.* [2]. This include observing the increase of song similarity (a measure defined by Tchernichovski, Nottebohm, Ho, *et al.* [14]) over the development of the bird, that the song similarity decrease over night, but that this decrease is positively correlated with the song restitution at the end of the learning. To do that, we will use the same statistical test that Derégnaucourt, Mitra, Fehér, *et al.* [2] have done on the same set of features. We will therefore use

Then, we expect our model to yield at first protosyllables, then syllables and go into the crystallized song.

Finally, we want that our algorithm identify the same amount of GTE per second that are identified by Boari, Perl, Amador, *et al.* [4].

look  
at how  
I can  
evaluate  
this by  
cluster-  
ing

## Final Evaluation

The goal of building this model is being able to build new hypotheses that can be tested in behavioral or neurobiological experiment on the zebra finch. Once the model is fully working and respond to our expectations in reproducing the literature, we will study its behaviors so as to design new hypotheses.

## References

- [1] W.-c. Liu, T. J. Gardner, and F. Nottebohm, “Juvenile zebra finches can use multiple strategies to learn the same song,” *Proceedings of the National Academy of Sciences*, vol. 101, no. 52, pp. 18 177–18 182, 2004. [Online]. Available: <http://www.pnas.org/content/101/52/18177.short> (visited on 10/20/2016).
- [2] S. Derégnaucourt, P. P. Mitra, O. Fehér, C. Pytte, and O. Tchernichovski, “How sleep affects the developmental learning of bird song,” *Nature*, vol. 433, no. 7027, pp. 710–716, Feb. 17, 2005, ISSN: 0028-0836, 1476-4679. DOI: [10.1038/nature03275](https://doi.org/10.1038/nature03275). [Online]. Available: <http://www.nature.com/doifinder/10.1038/nature03275> (visited on 09/21/2016).
- [3] A. S. Dave, “Song replay during sleep and computational rules for sensorimotor vocal learning,” *Science*, vol. 290, no. 5492, pp. 812–816, Oct. 27, 2000, ISSN: 00368075, 10959203. DOI: [10.1126/science.290.5492.812](https://doi.org/10.1126/science.290.5492.812). [Online]. Available: <http://www.sciencemag.org/cgi/doi/10.1126/science.290.5492.812> (visited on 09/21/2016).
- [4] S. Boari, Y. S. Perl, A. Amador, D. Margoliash, and G. B. Mindlin, “Automatic reconstruction of physiological gestures used in a model of birdsong production,” *Journal of Neurophysiology*, vol. 114, no. 5, pp. 2912–2922, Nov. 15, 2015, ISSN: 0022-3077, 1522-1598. DOI: [10.1152/jn.00385.2015](https://doi.org/10.1152/jn.00385.2015). [Online]. Available: <http://jn.physiology.org/content/114/5/2912> (visited on 12/08/2016).
- [5] A. Amador, Y. S. Perl, G. B. Mindlin, and D. Margoliash, “Elemental gesture dynamics are encoded by song premotor cortical neurons,” *Nature*, vol. 495, no. 7439, pp. 59–64, Feb. 27, 2013, ISSN: 0028-0836, 1476-4687. DOI: [10.1038/nature11967](https://doi.org/10.1038/nature11967). [Online]. Available: <http://www.nature.com/doifinder/10.1038/nature11967> (visited on 09/20/2016).

- [6] A. Amador and G. B. Mindlin, “Low dimensional dynamics in birdsong production,” *The European Physical Journal B*, vol. 87, no. 12, Dec. 2014, ISSN: 1434-6028, 1434-6036. DOI: [10.1140/epjb/e2014-50566-5](https://doi.org/10.1140/epjb/e2014-50566-5). [Online]. Available: <http://link.springer.com/10.1140/epjb/e2014-50566-5> (visited on 11/03/2016).
- [7] G. F. Lynch, T. S. Okubo, A. Hanuschkin, R. H. Hahnloser, and M. S. Fee, “Rhythmic continuous-time coding in the songbird analog of vocal motor cortex,” *Neuron*, vol. 90, no. 4, pp. 877–892, May 2016, ISSN: 08966273. DOI: [10.1016/j.neuron.2016.04.021](https://doi.org/10.1016/j.neuron.2016.04.021). [Online]. Available: <http://linkinghub.elsevier.com/retrieve/pii/S0896627316301088> (visited on 12/08/2016).
- [8] M. A. Picardo, J. Merel, K. A. Katlowitz, D. Vallentin, D. E. Okobi, S. E. Benezra, R. C. Clary, E. A. Pnevmatikakis, L. Paninski, and M. A. Long, “Population-level representation of a temporal sequence underlying song production in the zebra finch,” *Neuron*, vol. 90, no. 4, pp. 866–876, May 2016, ISSN: 08966273. DOI: [10.1016/j.neuron.2016.02.016](https://doi.org/10.1016/j.neuron.2016.02.016). [Online]. Available: <http://linkinghub.elsevier.com/retrieve/pii/S0896627316001094> (visited on 12/01/2016).
- [9] A. P. Dempster, N. M. Laird, and D. B. Rubin, “Maximum likelihood from incomplete data via the EM algorithm,” *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 39, no. 1, pp. 1–38, 1977, ISSN: 00359246. [Online]. Available: <http://www.jstor.org/stable/2984875>.
- [10] A. Baranes and P.-Y. Oudeyer, “Active learning of inverse models with intrinsically motivated goal exploration in robots,” *Robotics and Autonomous Systems*, vol. 61, no. 1, pp. 49–73, Jan. 2013, ISSN: 09218890. DOI: [10.1016/j.robot.2012.05.008](https://doi.org/10.1016/j.robot.2012.05.008). [Online]. Available: <http://linkinghub.elsevier.com/retrieve/pii/S0921889012000644> (visited on 12/15/2016).
- [11] C.-H. Chou, P.-H. Liu, and B. Cai, “On the studies of syllable segmentation and improving MFCCs for automatic birdsong recognition,” *IEEE*, Dec. 2008, pp. 745–750. DOI: [10.1109/APSCC.2008.6](https://doi.org/10.1109/APSCC.2008.6). [Online]. Available: <http://ieeexplore.ieee.org/document/4780764/> (visited on 10/06/2016).
- [12] L. Muda, M. Begam, and I. Elamvazuthi, “Voice recognition algorithms using mel frequency cepstral coefficient (MFCC) and dynamic time warping (DTW) techniques,” *ARXIV preprint arXiv:1003.4083*, 2010. [Online]. Available: <http://arxiv.org/abs/1003.4083> (visited on 11/24/2016).
- [13] C. Moulin-Frier, P. Rouanet, and P.-Y. Oudeyer, “Explauto: An open-source python library to study autonomous exploration in developmental robotics,” presented at the ICDL-Epirob - International Conference on Development and Learning, Epirob, Oct. 13, 2014. [Online]. Available: <https://hal.inria.fr/hal-01061708/document> (visited on 12/15/2016).
- [14] O. Tchernichovski, F. Nottebohm, C. E. Ho, B. Pesaran, and P. P. Mitra, “A procedure for an automated measurement of song similarity,” *Animal Behaviour*, vol. 59, no. 6, pp. 1167–1176, Jun. 2000, ISSN: 00033472. DOI: [10.1006/anbe.1999.1416](https://doi.org/10.1006/anbe.1999.1416). [Online]. Available: <http://linkinghub.elsevier.com/retrieve/pii/S0003347299914161> (visited on 09/29/2016).