INTERNSHIP POSITIONS AT THE ECOLE NORMALE SUPERIEURE, PARIS

The Institute for Cognitive Studies at the Ecole Normale Supérieure in Paris is proposing research internships to students with an engineering/maths/computer science background in the *Cognitive Machine Learning* team.

The general aim of this project is to understand how babies spontaneously learn their first language by applying a *'reverse engineering*' approach, i.e., by constructing an artificial language learner that mimics the learning stages of the infant.

The internship will focus on one specific subproblem, for instance: how do infants extract words from speech, how do they construct phoneme categories, how do they figure out the meaning of words? (see detailed list below). To address this problem, the student will apply *weakly supervised* or *unsupervised*, *bio-inspired machine learning* algorithms, to large corpora of child-adult verbal interactions in several languages and compare the results with behavioral and/or neural recording data. In particular, the student will work with tools selected from:

- signal processing (speech, video, brain imaging features)
- deep neural networks (optimized on GPUs)
- sparse dictionary methods
- hierarchical non parametric Bayesian models
- other tools from Natural Language Processing (Finite State Transducers, MaxEnt models, parsers, etc)

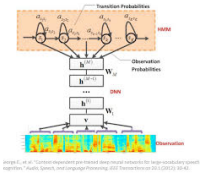
The student will work in a multidisciplinary team composed of researchers with various backgrounds (neuroscience, psycholinguistics, machine learning, etc) located at the Ecole Normale Supérieure in the quartier latin in Paris, and will have access to high performance computing resources (CPU/GPU cluster), large language databases, and cutting edge expertise in the cognitive (neuro)science of language as well as machine learning algorithms for speech and language applications. Some of the projects will involve a collaboration with other teams in France (INRIA) or abroad (J. Hopkins, MIT, Facebook AI Research, Google DeepMind, etc).

The student will ideally combine:

- a strong background in statistical modeling or linear algebra,
- knowledge of scientific computer programming (Matlab, python, etc).
- a strong interest in cognition and/or language,
- enthusiasm for interdisciplinary and team-based research,

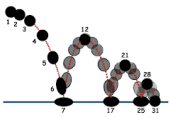
Candidates should send a CV, paragraph of motivation, contact information of one referee to *syntheticlearner@gmail.com*. Women are encouraged to apply. Further information about the project can be found at: http://www.syntheticlearner.net

Examples of possible internships



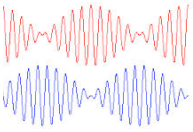
- Deep language learning from scratch

Deep Neural Networks (DNNs) have recently broken ground on state-of-the-art in several areas (image recognition, speech recognition, etc.)[1,2]. However, these algorithms depend on large human-annotated datasets. Yet, infants spontaneously achieve similar performance without direct supervision; the internship explores various ideas to 'de-supervise' deep learning using side information, loss functions or architectures inspired by research in human infants[3].



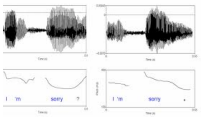
- **Learning the laws of physics with a deep recurrent network**

Recurrent networks can be used to learn regularities in video or audio sequences[4,5]. This internship will use a game engine to learn the underlying physical regularities of interactions between macroscopic objects and compare it to results of infant's perception of possible versus impossible events[6,7]. It will be conducted in collaboration with Facebook AI Research.



- **Time invariance in speech perception**.

Speech perception is invariant with respect to large variations in speech rate[8,9]. How is this achieved? The internship will explore time normalization using various computational architectures for speech recognition (convolutional coding, networks of oscillators, etc.) and compare the results to human data[10].



- **The role of prosody in language bootstrapping.**

Speech prosody is the 'melody' and 'rhythm' of language, and infants are very sensitive to it. We think that it provides bootstrapping into linguistic structures at many levels (lexical, grammatical)[11]. The internship will explore this using a variety of speech technology techniques (signal processing, spoken term discovery, word segmentation, etc.)[12].

- Rules and meaning

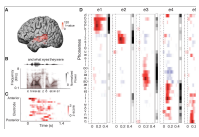
The human language faculty is unique in its ability to combine a finite number of categories to express infinitely varied meanings[13]. The internship addresses how the basic constituents of language (categories and rules) could be learned during infancy focusing on two ideas: extracting proto-categories and rules from the sensory inputs using clustering or sparse coding techniques[14], and using mutual constraints linking the different levels of linguistic structures[15].

- Multimodal language learning

At four months of age, infants recognize a few very common words (their names, mommy, daddy, etc)[16], even though they are unable to produce them. This internship tests whether multimodal DNNs can simultaneously learn words and their approximate meaning on a parallel dataset of audio and video tracks[17,18]. This internship will be conducted in collaboration with Microsoft Research at Redmond, USA.

- Massive baby home data collection

Big baby data is essential to uncover the mysteries of early language acquisition[19]. Here, we develop dense data recording in baby's homes using arrays of audio/3D video sensors[20], as well as toy-based evaluation of preverbal infant language acquisition, and we analyze the data in relation to computational models with unsupervised algorithms.



- Cracking the neural code for speech

How does the brain encode speech sounds? Progress in neuroimaging (ECoG, intracerebral electrical recording, etc) have resulted in a flow of data, both in human and animals[21,22]. The internship will apply neural decoding methods and apply to neural data and data generated from deep neural architectures[2] to explore hypotheses about the neural code for speech.

This list is not limitative
Visit us and discuss about other possible internships!

Bibliography

[1] Hinton, G., Deng, L., … Kingsbury, B. (2012). Deep Neural Networks for Acoustic Modeling in Speech Recognition: The Shared Views of Four Research Groups. *IEEE Signal Processing Magazine*, ***29*(6)**, 82–97.

[2] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems (pp. 1097-1105).

[3] Synnaeve, G., Schatz, T. & Dupoux, E. (2014). Phonetics embedding learning with side information. In *IEEE: SLT*.

[4] Srivastava, N., Mansimov, E., & Salakhutdinov, R. (2015). Unsupervised learning of video representations using lstms. arXiv Preprint arXiv:1502.04681..

[5] Sutskever, I., Vinyals, O., & Le, Q. V. (2014). Sequence to sequence learning with neural networks. In Advances in neural information processing systems (pp. 3104–3112).

[6] Franz, A., & Triesch, J. (2010). A unified computational model of the development of object unity, object permanence, and occluded object trajectory perception. *Infant Behavior and Development*, *33*(4), 635-653.

[7] Carey, S. (2009). *The origin of concepts*. Oxford University Press.

[8] Miller, J. L., & Baer, T. (1983). Some effects of speaking rate on the production of/b/and/w. *JASA*, **73(5)**, 1751-1755.

[9] Dupoux, E., & Green, K. (1997). Perceptual adjustment to highly compressed speech: effects of talker and rate changes. *Journal of Experimental Psychology: Human Perception and Performance*, **23(3)**, 914.

[10] Ghitza, O. (2014). Behavioral evidence for the role of cortical theta oscillations in determining auditory channel capacity for speech. *Frontiers in Psychology*, **5**

[11] Christophe, A., Guasti, M. T., Nespor, M., Dupoux, E., & van Ooyen, B. (1997). Reflections on prosodic bootstrapping: its role for lexical and syntactic acquisition. Language and Cognitive Processes, 12, 585-612

[12] Ludusan, B., Gravier, G. & Dupoux, E. (2014). Incorporating Prosodic Boundaries in Unsupervised Term Discovery. In *Speech Prosody-2014*.

[13] Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The faculty of language: What is it, who has it, and how did it evolve?. *Science*, **298(5598)**, 1569-1579.

[14] Christodoulopoulos, C., Goldwater, S., & Steedman, M. (2010, October). Two Decades of Unsupervised POS induction: How far have we come?. *In Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing* (pp. 575-584).

[15] Fourtassi, A., Dunbar, E. & Dupoux, E. (2014). Self Consistency as an Inductive Bias in Early Language Acquisition. In *Proceedings of Cog Sci*

[16] Bergelson, E., & Swingley, D. (Feb. 2012). At 6 to 9 months, human infants know the meanings of many common nouns. Proceedings of the National Academy of Sciences of the USA, 109, 3253-3258.

[17] Chen, X., & Zitnick, C. L. (2014). Learning a recurrent visual representation for image caption generation. *arXiv preprint arXiv:1411.5654*.

[18] Park, A. S., & Glass, J. R. (2008). Unsupervised Pattern Discovery in Speech. *IEEE Transactions on Audio, Speech, and Language Processing,* **16(1)**, 186–197.

[19] http://www.media.mit.edu/cogmac/projects/hsp.html

[20]http://www.microsoft.com/en-us/kinectforwindows/develop/

[21]. Mesgarani, N., Cheung, C., Johnson, K., & Chang, E. F. (2014). Phonetic Feature Encoding in Human Superior Temporal Gyrus. *Science*, **343(6174)**, 1006–1010.

[22] Mesgarani, N., David, S. V., Fritz, J. B., & Shamma, S. A. (2008). Phoneme representation and classification in primary auditory cortex. *The Journal of the Acoustical Society of America*, ***123*(2)**, 899.