Activity Report 2017

Team COML

Cognitive Machine Learning

Inria teams are typically groups of researchers working on the definition of a common project, and objectives, with the goal to arrive at the creation of a project-team. Such project-teams may include other partners (universities or research institutions).

RESEARCH CENTER
Paris

THEME
Language, Speech and Audio

Table of contents

Team COML

Creation of the Team: 2017 May 04

Keywords:

Computer Science and Digital Science:

A3.4.2. - Unsupervised learning
A3.4.5. - Bayesian methods
A3.4.6. - Neural networks
A3.4.8. - Deep learning
A5.7. - Audio modeling and processing
A5.7.1. - Sound
A5.7.3. - Speech
A5.7.4. - Analysis
A5.8. - Natural language processing
A5.9. - Signal processing
A5.9.1. - Sampling, acquisition
A5.9.2. - Estimation, modeling
A5.9.3. - Reconstruction, enhancement
A5.9.4. - Signal processing over graphs
A5.9.5. - Sparsity-aware processing
A5.9.6. - Optimization tools
A6.3.3. - Data processing
A9.2. - Machine learning
A9.3. - Signal analysis
A9.4. - Natural language processing
A9.6. - Decision support
A9.7. - AI algorithmics

Other Research Topics and Application Domains:

B1.2. - Neuroscience and cognitive science
B1.2.2. - Cognitive science

1. Personnel

Faculty Member

Emmanuel Dupoux [Team leader, Ecole Normale Supérieure Paris, Professor, from May 2017, HDR]

Post-Doctoral Fellow

Bogdan Ludusan [CNRS, until Oct 2017]

PhD Students

Maria Julia Carbajal [Ecole Normale Supérieure Paris]
Adriana Carolina Guevara Rukoz [Ecole Normale Supérieure Paris]
Neil Zeghidour [CIFRE Facebook; Ecole Normale Supérieure Paris]
Rahma Chaabouni [CIFRE Facebook; Ecole Normale Supérieure Paris, from Sep 2017]
Ronan Riochet [Ecole Normale Supérieure Paris, from Sep 2017; part time with the WILLOW team]

Elin Larsen [Ecole Normale Supérieure Paris, from Oct 2017]

Technical staff

Mathieu Bernard [Inria, from Nov 2017]
Julien Karadayi [Ecole Normale Supérieure]
Xuan Nga Cao [Ecole Normale Supérieure]

Administrative Assistants

Catherine Urban [Ecole Normale Supérieure]
Chantal Chazelas [Inria]

2. Overall Objectives

2.1. Overall Objectives

Brain-inspired machine learning algorithms combined with big data have recently reached spectacular results, equalling or beating humans on specific high level tasks (e.g. the game of go). However, there are still a lot of domains in which even humans infants outperform machines: unsupervised learning of rules and language, common sense reasoning, and more generally, cognitive flexibility (the ability to quickly transfer competence from one domain to another one).

The aim of the Cognitive Computing team is to *reverse engineer* such human abilities, i.e., to construct effective and scalable algorithms which perform as well (or better) than humans, when provided with similar data, study their mathematical and algorithmic properties and test their empirical validity as models of humans by comparing their output with behavioral and neuroscientific data. The expected results are more adaptable and autonomous machine learning algorithm for complex tasks, and quantitative models of cognitive processes which can used to predict human developmental and processing data. Most of the work is focused on speech and language and common sense reasoning.

3. Research Program

3.1. Background

In recent years, Artificial Intelligence (AI) has achieved important landmarks in matching or surpassing human level performance on a number of high level tasks (playing chess and go, driving cars, categorizing picture, etc., [72], [75], [80], [71], [77]). These strong advances were obtained by deploying on large amounts of data, massively parallel learning architectures with simple brain-inspired 'neuronal' elements. However, humans brains still outperform machines in several key areas (language, social interactions, common sense reasoning, motor skills), and are more flexible : Whereas machines require extensive expert knowledge and massive training for each particular application, humans learn autonomously over several time scales: over the developmental scale (months), humans infants acquire cognitive skills with noisy data and little or no expert feedback (weakly/unsupervised learning)[15]; over the short time scale (minutes, seconds), humans combine previously acquired skills to solve new tasks and apply rules systematically to draw inferences on the basis of extremely scarce data (learning to learn, domain adaptation, one- or zero-shot learning) [74].

The general aim of CoML, following the roadmap described in [15], is to bridge the gap in cognitive flexibility between humans and machines learning in language processing and common sense reasoning. We conduct work in three areas: weakly supervised and unsupervised algorithms, datasets and benchmarks, and machine intelligence evaluation.

3.2. Weakly/Unsupervised Learning

Much of standard machine learning is construed as regression or classification problems (mapping input data to expert-provided labels). Human infants rarely learn in this fashion, at least before going to school: they learn language, social cognition, and common sense autonomously (without expert labels) and when adults provide feedback, it is ambiguous and noisy and cannot be taken as a gold standard. Modeling or mimicking such achievement requires deploying unsupervised or weakly supervised algorithms which are less well known than their supervised counterparts.

We take inspiration from infant's landmarks during their first years of life: they are able to learn acoustic models, a lexicon, and susbtantive elements of language models and world models from raw sensory inputs. Building on previous work [21], [40], [54], we use DNN and Bayesian architectures to model the emergence of linguistic representations without supervision. Our focus is to establish how the labels in supervised settings can be replaced by weaker signals coming either from multi-modal input or from hierarchically organised linguistic levels.

At the level of phonetic representations, we study how cross-modal information (lips and self feedback from articulation) can supplement top-down lexical information in a weakly supervised setting. We use siamese architectures or Deep CCA algorithms to combine the different views. We study how an attentional framework and uncertainty estimation can flexibly combine these informations in order to adapt to situations where one view is selectively degraded.

At the level of lexical representations, we study how audio/visual parallel information (ie. descriptions of images or activities) can help in segmenting and clustering word forms, and vice versa, help in deriving useful visual features. To achieve this, we will use architectures deployed in image captioning or sequence to sequence translation [78].

At the level of semantic and conceptual representations, we study how it is possible to learn elements of the laws of physics through the observation of videos (object permanence, solidity, spatio-temporal continuity, inertia, etc.), and how objects and relations between objects are mapped onto language.

3.3. Evaluating Machine Intelligence

Increasingly, complicated machine learning systems are being incorporated into real-life applications (e.g. self-driving cars, personal assistants), even though they cannot be formally verified, guaranteed statistically, nor even explained. In these cases, a well defined *empirical approach* to evaluation can offer interesting insights into the functioning and offer some control over these algorithms.

Several approaches exist to evaluate the 'cognitive' abilities of machines, from the subjective comparison of human and machine performance [79] to application-specific metrics (e.g., in speech, word error rate). A recent idea consist in evaluating an AI system in terms of it's *abilities* [73] , i.e., functional components within a more global cognitive architecture [76]. Psychophysical testing can offer batteries of tests using simple tasks that are easy to understand by humans or animals (e.g, judging whether two stimuli are same or different, or judging whether one stimulus is 'typical') which can be made selective to a specific component and to rare but difficult or adversarial cases. Evaluations of learning rate, domain adaptation and transfer learning are simple applications of these measures. Psychophysically inspired tests have been proposed for unsupervised speech and language learning [46], [28].

3.4. Documenting human learning

Infants learn their first language in a spontaneous fashion, across a lot of variation in amount of speech and the nature of the infant/adult interaction. In some linguistic communities, adults barely address infants until they can themselves speak. Despite these large variations in quantity and content, language learning proceeds at similar paces. Documenting such resilience is an essential step in understanding the nature of the learning algorithms used by human infants. Hence, we propose to collect and/or analyse large datasets of inputs to infants and correlate this with outcome measure (phonetic learning, vocabulary growth, syntactic learning, etc.).

4. Application Domains

4.1. Speech processing for underresourced languages

We plan to apply our algorithms for the unsupervised discovery of speech units to problems relevant to language documentation and the construction of speech processing pipelines for underresourced languages.

4.2. Tools for the analysis of naturalistic speech corpora

Daylong recordings of speech in the wild gives rise a to number of specific analysis difficulties. We plan to use our expertise in speech processing to develop tools for performing signal processing and helping annotation of such resources for the purpose of phonetic or linguistic analysis.

5. Highlights of the Year

5.1. Highlights of the Year

5.1.1. Awards

BEST PAPER AWARD:

[67]
B. LUDUSAN, R. MAZUKA, M. BERNARD, A. CRISTIA, E. DUPOUX. *The Role of Prosody and Speech Register in Word Segmentation: A Computational Modelling Perspective*, in "Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)", Vancouver, Canada, Association for Computational Linguistics, July 2017 [*DOI :* 10.18653/v1/P17-2028], https://hal.inria.fr/hal-01687451

6. New Software and Platforms

6.1. abkhazia

KEYWORDS: Speech recognition - Speech-text alignment
FUNCTIONAL DESCRIPTION: The Abkhazia sofware makes it easy to obtain simple baselines for supervised ASR (using Kaldi) and ABX tasks (using ABXpy) on the large corpora of speech recordings typically used in speech engineering, linguistics or cognitive science research.

- Contact: Emmanuel Dupoux
- URL: https://github.com/bootphon/abkhazia

6.2. TDE

Term Discovery Evaluation
KEYWORDS: NLP - Speech recognition - Speech
SCIENTIFIC DESCRIPTION: This toolbox allows the user to judge of the quality of a word discovery algorithm. It evaluates the algorithms on these criteria : - Boundary : efficiency of the algorithm to found the actual boundaries of the words - Group : efficiency of the algorithm to group similar words - Token/Type: efficiency of the algorithm to find all words from the corpus (types), and to find all occurences (token) of these words. - NED : Mean of the edit distance across all the word pairs found by the algorithm - Coverage : efficiency of the algorithm to find every discoverable phone in the corpus

FUNCTIONAL DESCRIPTION: Toolbox to evaluate algorithms that segment speech into words. It allows the user to evaluate the efficiency of algorithms to segment speech into words, and create clusters of similar words.

- Contact: Emmanuel Dupoux
- URL : https://github.com/bootphon/TDE

6.3. ABXpy

KEYWORDS: Evaluation - Speech recognition - Machine learning
FUNCTIONAL DESCRIPTION: The ABX package gives a performance score to speech recognition systems by measuring their capacity to discriminate linguistic contrasts (accents, phonemes, speakers, etc...)

- Contact: Emmanuel Dupoux
- URL : https://github.com/bootphon/ABXpy

6.4. h5features

KEYWORD: File format
FUNCTIONAL DESCRIPTION: The h5features python package provides easy to use and efficient storage of large features data on the HDF5 binary file format.

- Contact: Emmanuel Dupoux
- URL : https://github.com/bootphon/h5features

7. New Results

7.1. Development of cognitively inspired algorithms

Speech and language processing in humans infants and adults is particularly efficient. We use these as sources of inspiration for developing novel machine learning and speech technology algorithms. In this area, our results are as follows:

- Recent works have explored deep architectures for learning multimodal speech representation (e.g. audio and images, articulation and audio) in a supervised way. In [63], we investigate the role of combining different speech modalities, i.e. audio and visual information representing the lips' movements, in a weakly-supervised way using Siamese networks and lexical same-different side information. In particular, we ask whether one modality can benefit from the other to provide a richer representation for phone recognition in a weakly supervised setting. We introduce mono-task and multi-task methods for merging speech and visual modalities for phone recognition. The mono-task learning consists in applying a Siamese network on the concatenation of the two modalities, while the multi-task learning receives several different combinations of modalities at train time. We show that multi-task learning enhances discriminability for visual and multimodal inputs while minimally impacting auditory inputs. Furthermore, we present a qualitative analysis of the obtained phone embeddings, and show that cross-modal visual input can improve the discriminability of phonetic features which are visually discernable (rounding, open/close, labial place of articulation), resulting in representations that are closer to abstract linguistic features than those based on audio only.
- In [67] , we explore the role of speech register and prosody for the task of word segmentation. Since these two factors are thought to play an important role in early language acquisition, we aim to quantify their contribution for this task. We study a Japanese corpus containing both infant- and adult-directed speech and we apply four different word segmentation models, with and without knowledge of prosodic boundaries. The results showed that the difference between registers is smaller than previously reported and that prosodic boundary information helps more adult- than infant-directed speech.

- Phonemic segmentation of speech is a critical step of speech recognition systems. In [68], we propose a novel unsupervised algorithm based on sequence prediction mod- els such as Markov chains and recurrent neural networks. Our approach consists in analyzing the error profile of a model trained to predict speech features frame- by-frame. Specifically, we try to learn the dynamics of speech in the MFCC space and hypothesize boundaries from lo- cal maxima in the prediction error. We evaluate our system on the TIMIT dataset, with improvements over similar methods.
- In [70], we describe a new challenge aimed at discovering subword and word units from raw speech. This challenge is the follow-up to the Zero Resource Speech Challenge 2015. It aims at constructing systems that generalize across languages and adapt to new speakers. The design features and evaluation metrics of the challenge are presented and the results of seventeen models are discussed.

7.2. Evaluation of AI algorithms

Machine learning algorithms are typically evaluated in terms of end-to-end tasks, but it is very often difficult to get a grasp of how they achieve these tasks, what could be their break point, and more generally, how they would compare to the algorithms used by humans to do the same tasks. This is especially true of Deep Learning systems which are particularly opaque. The team develops evaluation methods based on psycholinguistic/linguistic criteria, and deploy them for systematic comparison of systems.

- What is the information captured by neural network models of language? In [66], we address this question in the case of character-level recurrent neural language models. These models do not have explicit word representations; do they acquire implicit ones? We assess the lexical capacity of a network using the lexical decision task common in psycholinguistics: the system is required to decide whether or not a string of characters forms a word. We explore how accuracy on this task is affected by the architecture of the network, focusing on cell type (LSTM vs. SRN), depth and width. We also compare these architectural properties to a simple count of the parameters of the network. The overall number of parameters in the network turns out to be the most important predictor of accuracy; in particular, there is little evidence that deeper networks are beneficial for this task.

7.3. Learnability relevant descriptions of linguistic corpora

Evidently, infants are acquiring their language based on whatever linguistic input is available around them. The extent of variation that can be found across languages, cultures and socio-economic background provides strong constraints (lower bounds on data, higher bounds on noise, and variation and ambiguity) for language learning algorithms.

- In [60], we provide an estimation of how frequently, and from whom, children aged 0-11 years (Ns between 9 and 24) receive one-on-one verbal input among Tsimane forager-horticulturalists of lowland Bolivia. Analyses of systematic daytime behavioral observations reveal < 1 min per daylight hour is spent talking to children younger than 4 years of age, which is 4 times less than estimates for others present at the same time and place. Adults provide a majority of the input at 0–3 years of age but not afterward. When integrated with previous work, these results reveal large cross-cultural variation in the linguistic experiences provided to young children. Consideration of more diverse human populations is necessary to build generalizable theories of language acquisition.
- In [69], we provide a new measure of how the acoustic realizations of a given phonetic segment are affected by coarticulation with the preceding and following phonetic context. While coarticulation has been extensively studied using descriptive phonetic measurements, little is known about the functional impact of coarticulation for speech processing, and in particular, learnability. Here, we use DTW-based similarity defined on raw acoustic features and ABX scores to derive a measure of the effect of coarticulation on phonetic discriminability. This measure does not rely on defining segment-specific phonetic cues (formants, duration, etc.) and can be applied systematically and automatically to any segment in large scale corpora. We illustrate our method using stimuli in English and Japanese.

We replicate some well-known results, i.e., stronger anticipatory than perseveratory coarticulation and stronger coarticulation for lax/short vowels than for tense/long vowels. We then quantify for the first time the impact of coarticulation across different segment types (like vowels and consonants).

7.4. Test of the psychological validity of AI algorithms.

In this section, we focus on the utilisation of machine learning algorithms of speech and language processing to derive testable quantitative predictions in humans (adults or infants).

- In [61] we aim to quantify the relative contributions of phonetic categories and acoustic detail on phonotactically induced perceptual vowel epenthesis in Japanese listeners. A vowel identification task tested whether a vowel was perceived within illegal consonant clusters and, if so, which vowel was heard. Cross-spliced stimuli were used in which vowel coarticulation present in the cluster did not match the quality of the flanking vowel. Two clusters were used, /hp/ and /kp/, the former containing larger amounts of resonances of the preceding vowel. While both flanking vowel and coarticulation influenced vowel quality, the influence of coarticulation was larger, especially for /hp/.
- In [64], we explore the well documented example of vowel epenthesis, a phenomenon in which non-existent vowels are hallucinated by listeners, for stimuli containingr illegal consonantal sequences. As reported in previous work, this occurs in Japanese (JP) and Brazilian Portuguese (BP), languages for which the 'default' epenthetic vowels are /u/ and /i/, respectively. In a perceptual experiment, we corroborate the finding that the quality of this illusory vowel is language-dependent, but also that this default choice can be overridden by coarticulatory information present on the consonant cluster. In a second step, we analyse recordings of JP and BP speakers producing 'epenthesized' versions of stimuli from the perceptual task. Results reveal that the default vowel corresponds to the vowel with the most reduced acoustic characteristics, also the one for which formants are acoustically closest to formant transitions present in consonantal clusters. Lastly, we model behavioural responses from the perceptual experiment with an exemplar model using dynamic time warping (DTW)-based similarity measures on MFCCs.
- A range of computational approaches have been used to model the discovery of word forms from continuous speech by infants. Typically, these algorithms are evaluated with respect to the ideal 'gold standard' word segmentation and lexicon. These metrics assess how well an algorithm matches the adult state, but may not reflect the intermediate states of the child's lexical development. In [65], we set up a new evaluation method based on the correlation between word frequency counts derived from the application of an algorithm onto a corpus of child-directed speech, and the proportion of infants knowing the words according to parental reports. We evaluate a representative set of 4 algorithms, applied to transcriptions of the Brent corpus, which have been phonologized using either phonemes or syllables as basic units. Results show remarkable variation in the extent to which these 8 algorithm-unit combinations predicted infant vocabulary, with some of these predictions surpassing those derived from the adult gold standard segmentation. We argue that infant vocabulary prediction provides a useful complement to traditional evaluation; for example, the best predictor model was also one of the worst in terms of segmentation score, and there was no clear relationship between token or boundary F-score and vocabulary prediction.
- A central assumption of most computational models of language acquisition is the reliance on statistical processes. This would predict that the frequency of particular sounds or contrasts in a given language should have a massive effect on perception. Surprisingly , this has not up to now been put to empirical test. In [62], we elucidated indicators of frequency-dependent perceptual attunement in the brain of 5–8-month-old Dutch infants. We tested the' discrimination of tokens containing a highly frequent [haet-he:t] and a highly infrequent [hYt-hø:t] native vowel contrast as well as a non-native [htˆ-hæt] vowel contrast in a behavioral visual habituation paradigm (Experiment 1). Infants discriminated both native contrasts similarly well, but did not discriminate the non-native contrast. We sought further evidence for subtle differences in the processing of the two native contrasts using

near-infrared spectroscopy and a within-participant design (Experiment 2). The neuroimaging data did not provide additional evidence that responses to native contrasts are modulated by frequency of exposure. These results suggest that even large differences in exposure to a native contrast may not directly translate to behavioral and neural indicators of perceptual attunement, raising the possibility that frequency of exposure does not influence improvements in discriminating native contrasts.

8. Bilateral Contracts and Grants with Industry

8.1. Bilateral Grants with Industry

- Grant from MSR (Zero Resources Challenge, 2017) - 5K€
- AWS Grant (Zero Resources Challenge, 2017) - 20K€

9. Partnerships and Cooperations

9.1. Regional Initiatives

Collaboration with the Willow Team:

- co-advising with J. Sivic and I. Laptev of a PhD student: Ronan Riochet.
- construction of a naive physics benchmark

9.2. National Initiatives

9.2.1. ANR

Transatlantic Platform "Digging into Data". Title: "Analysis of Children's Language Experiences Around the World. (ACLEW)"; (coordinating PI : M. Soderstrom; Leader of tools development and co-PI : E. Dupoux), (2017–2020. 5 countries; Total budget: 1.4M€)

9.3. European Initiatives

9.3.1. FP7 & H2020 Projects

ERC Advanced Grant (BOOTPHON, PI: E. Dupoux, Budget 2.4M€).

9.4. International Initiatives

9.4.1. Informal International Partners

- Johns Hopkins University, Baltimore, USA: S. Kudanpur, H. Hermanksy
- RIKEN Institute, Tokyo, Japan: R. Mazuka

9.5. International Research Visitors

9.5.1. Visits of International Scientists

Valentina Gliozzi (Professor, Univ. di Torino, Visiting Professor Spring 2017)

10. Dissemination

10.1. Promoting Scientific Activities

10.1.1. Scientific Events Organisation

10.1.1.1. General Chair, Scientific Chair

Zero Resource Challenge 2017, held as a special session of EEE ASRU 2017, Okinawa.

10.1.1.2. Member of the Organizing Committees

Executive committee of SIGMORPHON (Association for Computational Linguistics Special Interest Group, http://www.sigmorphon.org/).

10.1.2. Scientific Events Selection

10.1.2.1. Reviewer

Invited editor for international conferences: Interspeech, NIPS, ACL, etc. (around 5-10 papers per conferences, 2 conferences per year)

10.1.3. Journal

10.1.3.1. Member of the Editorial Boards

Member of the editorial board of: Mathématiques et Sciences Humaines, L'Année Psychologique, Frontiers in Psychology.

10.1.3.2. Reviewer - Reviewing Activities

Invited Reviewer for Frontiers in Psychology, Cognitive Science, Cognition, Transactions in Acoustics Signal Processing and Language, Speech Communication, etc. (around 4 papers per year)

10.1.4. Invited Talks

- Learning in Machines and Brains (CIFAR) invited talk, 2017, Paris.
- CBMM (Center for Brain Mind and Machine) Workshop on Speech representation, perception and recognition. Invited talk. Feb 02-03, 2017, MIT

10.1.5. Scientific Expertise

E. Dupoux is invited expert for ERC, ANR, and other granting agencies (around 2 per year).

10.1.6. Research Administration

Executive committee of the Foundation Cognition, the research programme IRIS-PSL "Sciences des Données et Données des Sciences", the industrial chair Almerys (2016-) and the collective organization DARCLE (www.darcle.org).

10.2. Teaching - Supervision - Juries

10.2.1. Teaching

Licence : E. Dupoux, "Introduction to the cognitive science of language", 8h, L2, PSL, France

Master : E. Dupoux, "Theoretical Cognitive Science: Connections and symbols", 8h, M1/M2, PSL,Paris 5, Paris France

Master : E. Dupoux, "Cognitive Engineering", 80h, M2, ITI-PSL, Paris France

Doctorat : E. Dupoux, "Computational models of cognitice development", 32 h, Séminaire EHESS, Paris France

10.2.2. Supervision

Six PhD theses are currently conducted in the team. Two are programmed to be defended in september 2018.

10.2.3. Juries

E. Dupoux participated in the PhD jury of Martin Felipe Perez-Guevara on novembre 29th, 2017 at UPMC (supervisor: C. Pallier).

10.3. Popularization

E. Dupoux talked in two general public conferences on speech recognition, one organised par France is AI (oct 17, 2017) and the other by Paris Sciences& Data (dec 7, 2017). The public was composed of entrepreneurs in machine learning. He also gave a training course in speech and language technology organised by the Institut de l'Ecole Normale (june 21) aimed at information technology professionals. He gave three interview on the limits of deep learning in general public outlets (Le Monde, La Recherche [1], Usine Nouvelle).

N. Zeghidour did a high level presentation of AI to 30 students and 8 teachers from Sciences-Po and Ecole 42 during the Policy Innovation Lab [2] following the visit of Facebook's COO. He has been a Mentor (technical advisor) for 9 European start-ups during a pitching event for the IBM Watson AI Xprize [3]. He animated a booth on AI at the Platform Meetup Paris of Facebook. He gave an interview to Libération included in a 100 page special issue on AI [4]. He taught a class on Speech Recognition as part of the Facebook AI Masterclass, a class for developers located at Ecole 42 and broadcasted in developer circles over 25 cities of Europe, Middle-Est and Africa with 850 people attending, followed by a live Q&A session with several of these cities over video-conference.

11. Bibliography

Major publications by the team in recent years

[1] C. BERGMANN, A. CRISTIA, E. DUPOUX. *Discriminability of sound contrasts in the face of speaker variation quantified*, in "Proceedings of the 38th Annual Conference of the Cognitive Science Society", 2016, pp. 1331-1336

[2] M. BUON, E. DUPOUX, P. JACOB, P. CHASTE, M. LEBOYER, T. ZALLA. *The role of causal and intentional reasoning in moral judgment in individuals with High Functioning Autism*, in "Journal of Autism and Developmental Disorders", 2013, vol. 43, n° 2, pp. 458-70 [*DOI :* 10.1016/J.COGNITION.2012.09.006]

[3] M. BUON, P. JACOB, E. LOISSEL, E. DUPOUX. *A non-mentalistic cause-based heuristic in human social evaluations*, in "Cognition", 2013, vol. 126, n° 2, pp. 149-155

[4] M. BUON, P. JACOB, S. MARGULES, I. BRUNET, M. DUTAT, D. CABROL, E. DUPOUX. *Friend or foe? Early social evaluation of human interactions*, in "PloS One", 2014, vol. 9, n° 2, e88612

[5] J. CARBAJAL, A. DAWUD, R. THIOLLIÈRE, E. DUPOUX. *The 'Language Filter' Hypothesis: Modeling Language Separation in Infants using I-vectors*, in "EPIROB 2016", 2016, pp. 195-201

[6] J. CARBAJAL, R. FÉR, E. DUPOUX. *Modeling language discrimination in infants using i-vector representations*, in "Proceedings of the 38th Annual Conference of the Cognitive Science Society", 2016, pp. 889-896

[1] http://www.larecherche.fr/lintelligence-artificielle-a-lassaut-des-labos
[2] http://www.sciencespo.fr/public/en/policy-lab/public-policy-incubator
[3] https://ai.xprize.org/
[4] http://www.liberation.fr/voyage-au-coeur-de-lIA/2017/12/20/en-kiosque-notre-hors-serie-voyage-au-coeur-de-l-ia_1617896

[7] L. CLERET DE LANGAVANT, C. JACQUEMOT, A.-C. BACHOUD-LÉVI, E. DUPOUX. *The second person in 'I'-'you'-'it' triadic interactions*, in "Behavioral and Brain Sciences", 2013, vol. 36, n° 416-417

[8] A. CRISTIA, E. DUPOUX, Y. HAKUNA, S. LLOYD-FOX, M. SCHUETZE, J. KIVITS, T. BERGVELT, M. VAN GELDER, L. FILIPPIN, S. CHARRON, Y. MINAGAWA-KAWAI. *An online database of infant functional Near InfraRed Spectroscopy studies: A community-augmented systematic review.*, in "PLoS One", 2013, vol. 8, n° 3, e58906

[9] A. CRISTIA, Y. MINAGAWA-KAWAI, N. EGOROVA, J. GERVAIN, L. FILIPPIN, D. CABROL, E. DUPOUX. *Neural correlates of infant dialect discrimination: A fNIRS study*, in "Developmental Science", 2014, vol. 17, n° 4, pp. 628-635

[10] A. CRISTIA, Y. MINAGAWA-KAWAI, I. VENDELIN, D. CABROL, E. DUPOUX. *Responses to vocalizations and auditory controls in the human newborn brain*, in "Plos One", 2014, vol. 9, n° 12, e115162

[11] E. DUNBAR, E. DUPOUX. *Geometric constraints on human speech sound inventories*, in "Frontiers in Psychology", 2016, vol. 7, n° 1061 [*DOI :* 10.3389/FPSYG.2016.01061]

[12] E. DUNBAR, G. SYNNAEVE, E. DUPOUX. *Quantitative methods for comparing featural representations*, in "ICPhS", 2015, paper number 1024

[13] E. DUPOUX. *Towards Quantitative Studies of Early Cognitive Development*, in "Autonomous Mental Development Technical Committee Newsletter", 2014, vol. 11, n° 1, pp. 10-11

[14] E. DUPOUX. *Category Learning in Songbirds: top-down effects are not unique to humans*, in "Current Biology", 2015, vol. 25, n° 16, pp. R718-R720

[15] E. DUPOUX. *Cognitive Science in the era of Artificial Intelligence: A roadmap for reverse-engineering the infant language-learner*, in "Cognition", 2018

[16] A. FOURTASSI, B. BOERSCHINGER, M. JOHNSON, E. DUPOUX. *WhyisEnglishsoeasytosegment*, in "Proceedings of the 4th Workshop on Cognitive Modeling and Computational Linguistics (CMCL 2013)", Sofia, Bulgaria, ACL, 2013, pp. 1-10

[17] A. FOURTASSI, E. DUNBAR, E. DUPOUX. *Self Consistency as an Inductive Bias in Early Language Acquisition*, in "Proceedings of the 36th Annual Meeting of the Cognitive Science Society", CogSci, 2014, pp. 469-474

[18] A. FOURTASSI, E. DUPOUX. *A corpus-based evaluation method for Distributional Semantic Models*, in "Proceedings of ACL-SRW 2013", Sofia, Bulgaria, ACL, 2013, pp. 165-171

[19] A. FOURTASSI, E. DUPOUX. *A Rudimentary Lexicon and Semantics Help Bootstrap Phoneme Acquisition*, in "Proceedings of the 18th Conference on Computational Natural Language Learning (CoNLL)", Baltimore, Maryland USA, Association for Computational Linguistics, June 2014, pp. 191-200 [*DOI :* 10.3115/V1/W14-1620]

[20] A. FOURTASSI, E. DUPOUX. *The role of word-word co-occurrence in word learning*, in "Proceedings of the 38th Annual Conference of the Cognitive Science Society", 2016, pp. 662-667

[21] A. FOURTASSI, T. SCHATZ, B. VARADARAJAN, E. DUPOUX. *Exploring the Relative Role of Bottom-up and Top-down Information in Phoneme Learning*, in "Proceedings of the 52nd Annual meeting of the ACL", Baltimore, Maryland, Association for Computational Linguistics, 2014, vol. 2, pp. 1-6 [*DOI :* 10.3115/v1/P14-2001]

[22] K. GVOZDIC, S. MOUTIER, E. DUPOUX, M. BUON. *Priming Children's Use of Intentions in Moral Judgement with Metacognitive Training*, in "Frontiers in Language Sciences", 2016, vol. 7, n° 190 [*DOI :* 10.3389/FPSYG.2016.00190]

[23] H. HERMANSKY, L. BURGET, J. COHEN, E. DUPOUX, N. FELDMAN, J. GODFREY, S. KHUDANPUR, M. MACIEJEWSKI, S. H. MALLIDI, A. MENON, T. OGAWA, V. PEDDINTI, R. ROSE, R. STERN, M. WIESNER, K. VESELY. *Towards machines that know when they do not know: Summary of work done at 2014 Frederick Jelinek memorial workshop in Prague*, in "ICASSP-2015 (IEEE International Conference on Acoustics Speech and Signal Processing)", Brisbane, Australia, 19-24 April 2015, pp. 5009-5013

[24] A. JANSEN, E. DUPOUX, S. GOLDWATER, M. JOHNSON, S. KHUDANPUR, K. CHURCH, N. FELDMAN, H. HERMANSKY, F. METZE, R. ROSE, M. SELTZER, P. CLARK, I. MCGRAW, B. VARADARAJAN, E. BENNETT, B. BORSCHINGER, J. CHIU, E. DUNBAR, A. FOURTASSI, D. HARWATH, C.-Y. LEE, K. LEVIN, A. NOROUZIAN, V. PEDDINTI, R. RICHARDSON, T. SCHATZ, S. THOMAS. *A summary of the 2012 JH CLSP Workshop on zero resource speech technologies and models of early language acquisition*, in "ICASSP-2013 (IEEE International Conference on Acoustics Speech and Signal Processing)", Vancouver, BC, Canada, IEEE, May 2013, pp. 8111-8115 [*DOI :* 10.1109/ICASSP.2013.6639245]

[25] M. JOHNSON, A. CHRISTOPHE, K. DEMUTH, E. DUPOUX. *Modelling function words improves unsupervised word segmentation*, in "Proceedings of the 52nd Annual meeting of the ACL", ACL, 2014, pp. 282–292 [*DOI :* 10.3115/v1/P14-1027]

[26] M. JOHNSON, J. PATER, R. STAUB, E. DUPOUX. *Sign constraints on feature weights improve a joint model of word segmentation and phonology*, in "NAACL HLT 2015", ACL, 2015, pp. 303-313 [*DOI :* 10.3115/v1/N15-1034]

[27] T. LINZEN, E. DUPOUX, Y. GOLDBERG. *Assessing the ability of LSTMs to learn syntax-sensitive dependencies*, in "Transactions of the Association for Computational Linguistics", 2016, vol. 4, pp. 521-535

[28] T. LINZEN, E. DUPOUX, B. SPECTOR. *Quantificational features in distributional word representations*, in "Proceedings of the Fifth Joint Conference on Lexical and Computational Semantics", 2016, pp. 1-11 [*DOI :* 10.18653/v1/S16-2001]

[29] B. LUDUSAN, A. CARANICA, H. CUCU, A. BUZO, C. BURILEANU, E. DUPOUX. *Exploring multi-language resources for unsupervised spoken term discovery*, in "Speech Technology and Human-Computer Dialogue (SpeD), 2015 International Conference on", 2015, pp. 1-6

[30] B. LUDUSAN, A. CRISTIA, A. MARTIN, R. MAZUKA, E. DUPOUX. *Learnability of prosodic boundaries: Is infant-directed speech easier?*, in "Journal of the Acoustical Society of America", 2016, vol. 140, n° 2, pp. 1239-1250

[31] B. LUDUSAN, E. DUPOUX. *Towards Low Resource Prosodic Boundary Detection*, in "Proceedings of International Workshop on Spoken Language Technologies for Under-resourced Languages (SLTU'14)", St Petersburg, Russia, SLTU, May 14-16 2014, pp. 231-237

[32] B. LUDUSAN, E. DUPOUX. *A multilingual study on intensity as a cue for marking prosodic boundaries*, in "ICPhS", 2015, e982

[33] B. LUDUSAN, E. DUPOUX. *Automatic syllable segmentation using broad phonetic class information*, in "SLTU-2016 Procedia Computer Science", 2016, vol. 81, pp. 101-106

[34] B. LUDUSAN, E. DUPOUX. *The role of prosodic boundaries in word discovery: Evidence from a computational model*, in "Journal of the Acoustical Society of America", 2016, vol. 140, n^{o} 1, EL1

[35] B. LUDUSAN, G. GRAVIER, E. DUPOUX. *Incorporating Prosodic Boundaries in Unsupervised Term Discovery*, in "Proceedings of Speech Prosody", Dublin, Ireland, 2014, vol. 7, pp. 939-943

[36] B. LUDUSAN, A. ORIGLIA, E. DUPOUX. *Rhythm-Based Syllabic Stress Learning without Labelled Data*, in "Proceedings of Statistical Language and Speech Processing -SLSP 2015", 2015, pp. 185-196

[37] B. LUDUSAN, A. SEIDL, E. DUPOUX, A. CRISTIA. *Motif discovery in infant- and adult-directed speech*, in "Proceedings of CogACLL2015", ACL, 2015, pp. 93-102 [*DOI :* 10.18653/V1/W15-2413]

[38] B. LUDUSAN, G. SYNNAEVE, E. DUPOUX. *Prosodic boundary information helps unsupervised word segmentation*, in "NAACL HLT 2015", ACL, 2015, pp. 953-963 [*DOI :* 10.3115/V1/N15-1096]

[39] B. LUDUSAN, M. VERSTEEGH, A. JANSEN, G. GRAVIER, X.-N. CAO, M. JOHNSON, E. DUPOUX. *Bridging the gap between speech technology and natural language processing: an evaluation toolbox for term discovery systems*, in "Proceedings of LREC 2014", Some place, LREC, 2014, pp. 560-567

[40] A. MARTIN, S. PEPERKAMP, E. DUPOUX. *Learning Phonemes with a Proto-lexicon*, in "Cognitive Science", 2013, vol. 37, pp. 103-124 [*DOI :* 10.1111/J.1551-6709.2012.01267.X]

[41] A. MARTIN, T. SCHATZ, M. VERSTEEGH, K. MIYAZAWA, R. MAZUKA, E. DUPOUX, A. CRISTIA. *Mothers speak less clearly to infants: A comprehensive test of the hyperarticulation hypothesis*, in "Psychological Science", 2015, vol. 26, n^{o} 3, pp. 341-347 [*DOI :* 10.1177/0956797614562453]

[42] E. MICHON, E. DUPOUX, A. CRISTIA. *Salient dimensions in implicit phonotactic learning*, in "INTERSPEECH-2015", 2015, pp. 2665-2669

[43] Y. MINAGAWA-KAWAI, A. CRISTIA, B. LONG, I. VENDELIN, Y. HAKUNO, M. DUTAT, L. FILIPPIN, D. CABROL, E. DUPOUX. *Insights on NIRS sensitivity from a cross-linguistic study on the emergence of phonological grammar*, in "Frontiers in Language Sciences", April 16 2013, vol. 4, n^{o} 170 [*DOI :* 10.3389/FPSYG.2013.00170]

[44] C. NGON, A. MARTIN, E. DUPOUX, D. CABROL, S. PEPERKAMP. *Nonwords, nonwords, nonwords: Evidence for a proto-lexicon during the first year of life*, in "Developmental Science", 2013, vol. 16, n^{o} 1, pp. 24-34 [*DOI :* 10.1111/J.1467-7687.2012.01189.X]

[45] T. OGAWA, S. H. MALLIDI, E. DUPOUX, J. COHEN, N. FELDMAN, H. HERMANSKY. *A new efficient measure for accuracy prediction and its application to multistream-based unsupervised adaptation*, in "ICPR", 2016

[46] T. SCHATZ, V. PEDDINTI, F. BACH, A. JANSEN, H. HYNEK, E. DUPOUX. *Evaluating speech features with the Minimal-Pair ABX task: Analysis of the classical MFC/PLP pipeline*, in "INTERSPEECH-2013", Lyon, France, International Speech Communication Association, 2013, pp. 1781-1785

[47] T. SCHATZ, V. PEDDINTI, X.-N. CAO, F. BACH, H. HYNEK, E. DUPOUX. *Evaluating speech features with the Minimal-Pair ABX task (II): Resistance to noise*, in "INTERSPEECH-2014", International Speech Communication Association, 2014, pp. 915-919

[48] G. SYNNAEVE, I. DAUTRICHE, B. BOERSCHINGER, M. JOHNSON, E. DUPOUX. *Unsupervised word segmentation in context*, in "Proceedings of 25th International Conference on Computational Linguistics (CoLing)", CoLing, 2014, pp. 2326-2334

[49] G. SYNNAEVE, E. DUPOUX. *In Depth Deep Beliefs Networks for Phone Recognition*, in "Poster presented in NIPS-2013", NIPS, 2013

[50] G. SYNNAEVE, E. DUPOUX. *Weakly Supervised Multi-Embeddings Learning of Acoustic Models*, in "ICLR Workshop", 2015, ArXiv 1412.6645 [cs.SD]

[51] G. SYNNAEVE, E. DUPOUX. *A temporal coherence loss function for learning unsupervised acoustic embeddings*, in "SLTU-2016 Procedia Computer Science", ISCA-ITRW, 2016, vol. 81, pp. 95-100

[52] G. SYNNAEVE, T. SCHATZ, E. DUPOUX. *Phonetics embedding learning with side information*, in "IEEE Spoken Language Technology Workshop", IEEE, 2014, pp. 106 - 111 [*DOI :* 10.1109/SLT.2014.7078558]

[53] G. SYNNAEVE, M. VERSTEEGH, E. DUPOUX. *Learning words from images and speech*, in "NIPS Workshop on Learning Semantics", Montreal, Canada, 2014

[54] R. THIOLLIÈRE, E. DUNBAR, G. SYNNAEVE, M. VERSTEEGH, E. DUPOUX. *A Hybrid Dynamic Time Warping-Deep Neural Network Architecture for Unsupervised Acoustic Modeling*, in "INTERSPEECH-2015", 2015, pp. 3179-3183

[55] M. VERSTEEGH, X. ANGUERA, A. JANSEN, E. DUPOUX. *The Zero Resource Speech Challenge 2015: Proposed Approaches and Results*, in "SLTU-2016 Procedia Computer Science", ISCA-ITRW, 2016, vol. 81, pp. 67-72

[56] M. VERSTEEGH, R. THIOLLIÈRE, T. SCHATZ, X.-N. CAO, X. ANGUERA, A. JANSEN, E. DUPOUX. *The Zero Resource Speech Challenge 2015*, in "INTERSPEECH-2015", 2015, pp. 3169-3173

[57] N. ZEGHIDOUR, G. SYNNAEVE, N. USUNIER, E. DUPOUX. *Joint Learning of Speaker and Phonetic Similarities with Siamese Networks*, in "INTERSPEECH-2016", ISCA, 2016, pp. 1295-1299

[58] N. ZEGHIDOUR, G. SYNNAEVE, M. VERSTEEGH, E. DUPOUX. *A Deep Scattering Spectrum - Deep Siamese Network Pipeline For Unsupervised Acoustic Modeling*, in "ICASSP-2016", IEEE, 2016, pp. 4965-4969

[59] R. DE DIEGO-BALAGUER, C. SCHRAMM, I. REBEIX, E. DUPOUX, A. DURR, A. BRICE, P. CHARLES, L. CLERET DE LANGAVANT, K. YOUSSOV, C. VERNY, V. DAMOTTE, J.-P. AZULAY, C. GOIZET, C. SIMONIN, C. TRANCHANT, P. MAISON, A. RIALLAND, D. SCHMITZ, C. JACQUEMOT, B. FONTAINE,

A.-C. BACHOUD-LÉVI. *COMT Val158Met Polymorphism Modulates Huntington's Disease Progression*, in "Plos One", 2016, vol. 11, n^{o} 9, e0161106 [*DOI :* 10.1371/JOURNAL.PONE.0161106]

Publications of the year

Articles in International Peer-Reviewed Journals

[60] A. CRISTIA, E. DUPOUX, M. GURVEN, J. STIEGLITZ. *Child-Directed Speech Is Infrequent in a Forager-Farmer Population: A Time Allocation Study*, in "Child Development", 2017 [*DOI :* 10.1111/CDEV.12974], https://hal.inria.fr/hal-01687336

[61] A. GUEVARA-RUKOZ, I. LIN, M. MORII, Y. MINAGAWA-KAWAI, E. DUPOUX, S. PEPERKAMP. *Which epenthetic vowel? Phonetic categories versus acoustic detail in perceptual vowel epenthesis*, in "Journal of the Acoustical Society of America", August 2017, vol. 142, n^{o} 2, pp. EL211 - EL217 [*DOI :* 10.1121/1.4998138], https://hal.inria.fr/hal-01687489

[62] S. TSUJI, P. FIKKERT, Y. MINAGAWA-KAWAI, E. DUPOUX, L. FILIPPIN, M. VERSTEEGH, P. HAGOORT, A. CRISTIA. *The more, the better? Behavioral and neural correlates of frequent and infrequent vowel exposure*, in "Developmental Psychobiology", July 2017, vol. 59, n^{o} 5, pp. 603 - 612 [*DOI :* 10.1002/DEV.21534], https://hal.inria.fr/hal-01687403

International Conferences with Proceedings

[63] R. CHAABOUNI, E. DUNBAR, N. ZEGHIDOUR, E. DUPOUX. *Learning Weakly Supervised Multimodal Phoneme Embeddings*, in "Interspeech 2017", Stockholm, Sweden, ISCA, 2017 [*DOI :* 10.21437/INTERSPEECH.2017-1689], https://hal.inria.fr/hal-01687415

[64] A. GUEVARA-RUKOZ, E. PARLATO-OLIVEIRA, S. YU, Y. HIROSE, S. PEPERKAMP, E. DUPOUX. *Predicting Epenthetic Vowel Quality from Acoustics*, in "Interspeech 2017", Stockholm, Sweden, ISCA, 2017 [*DOI :* 10.21437/INTERSPEECH.2017-1735], https://hal.inria.fr/hal-01687378

[65] E. LARSEN, E. DUPOUX, A. CRISTIA. *Relating Unsupervised Word Segmentation to Reported Vocabulary Acquisition*, in "Interspeech 2017", Stockholm, Sweden, ISCA, 2017 [*DOI :* 10.21437/INTERSPEECH.2017-937], https://hal.inria.fr/hal-01687534

[66] G. LE GODAIS, T. LINZEN, E. DUPOUX. *Comparing Character-level Neural Language Models Using a Lexical Decision Task*, in "Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers", Valencia, Spain, Association for Computational Linguistics, April 2017 [*DOI :* 10.18653/V1/E17-2020], https://hal.inria.fr/hal-01687583

[67] *Best Paper*
B. LUDUSAN, R. MAZUKA, M. BERNARD, A. CRISTIA, E. DUPOUX. *The Role of Prosody and Speech Register in Word Segmentation: A Computational Modelling Perspective*, in "Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)", Vancouver, Canada, Association for Computational Linguistics, July 2017 [*DOI :* 10.18653/V1/P17-2028], https://hal.inria.fr/hal-01687451.

[68] M. PAUL, O. RASANEN, R. THIOLLIÈRE, E. DUPOUX. *Blind Phoneme Segmentation With Temporal Prediction Errors*, in "Proceedings of ACL 2017, Student Research Workshop", Vancouver, Canada, Association

for Computational Linguistics, July 2017, https://arxiv.org/abs/1608.00508 [*DOI :* 10.18653/v1/P17-3011], https://hal.inria.fr/hal-01687524

[69] T. SCHATZ, R. TURNBULL, F. BACH, E. DUPOUX. *A Quantitative Measure of the Impact of Coarticulation on Phone Discriminability*, in "Interspeech 2017", Stockholm, Sweden, ISCA, 2017 [*DOI :* 10.21437/INTERSPEECH.2017-1306], https://hal.inria.fr/hal-01687436

Books or Proceedings Editing

[70] E. DUNBAR, X.-N. CAO, J. BENJUMEA, J. KARADAYI, M. BERNARD, L. BESACIER, X. ANGUERA, E. DUPOUX (editors). *The Zero Resource Speech Challenge 2017*, 2017, https://arxiv.org/abs/1712.04313 - IEEE ASRU (Automatic Speech Recognition and Understanding) 2017. Okinawa, Japan, https://hal.inria.fr/hal-01687504

References in notes

[71] D. A. FERRUCCI. *Introduction to "this is watson"*, in "IBM Journal of Research and Development", 2012, vol. 56, n^{o} 3.4, pp. 1–1

[72] K. HE, X. ZHANG, S. REN, J. SUN. *Delving deep into rectifiers: Surpassing human-level performance on imagenet classification*, in "Proceedings of the IEEE International Conference on Computer Vision", 2015, pp. 1026–1034

[73] J. HERNÁNDEZ-ORALLO, F. MARTÍNEZ-PLUMED, U. SCHMID, M. SIEBERS, D. L. DOWE. *Computer models solving intelligence test problems: Progress and implications*, in "Artificial Intelligence", 2016, vol. 230, pp. 74–107

[74] B. M. LAKE, T. D. ULLMAN, J. B. TENENBAUM, S. J. GERSHMAN. *Building machines that learn and think like people*, in "arXiv preprint arXiv:1604.00289", 2016

[75] C. LU, X. TANG. *Surpassing human-level face verification performance on LFW with GaussianFace*, in "arXiv preprint arXiv:1404.3840", 2014

[76] S. T. MUELLER. *A partial implementation of the BICA cognitive decathlon using the Psychology Experiment Building Language (PEBL)*, in "International Journal of Machine Consciousness", 2010, vol. 2, n^{o} 02, pp. 273–288

[77] D. SILVER, A. HUANG, C. J. MADDISON, A. GUEZ, L. SIFRE, G. VAN DEN DRIESSCHE, J. SCHRITTWIESER, I. ANTONOGLOU, V. PANNEERSHELVAM, M. LANCTOT, S. DIELEMAN, D. GREWE, J. NHAM, N. KALCHBRENNER, I. SUTSKEVER, T. LILLICRAP, M. LEACH, K. KAVUKCUOGLU, T. GRAEPEL, D. HASSABIS. *Mastering the game of Go with deep neural networks and tree search*, in "Nature", 2016, vol. 529, n^{o} 7587, pp. 484–489

[78] I. SUTSKEVER, O. VINYALS, Q. V. LE. *Sequence to sequence learning with neural networks*, in "Advances in neural information processing systems", 2014, pp. 3104–3112

[79] A. M. TURING. *Computing machinery and intelligence*, in "Mind", 1950, vol. 59, n^{o} 236, pp. 433–460

[80] W. XIONG, J. DROPPO, X. HUANG, F. SEIDE, M. SELTZER, A. STOLCKE, D. YU, G. ZWEIG. *Achieving human parity in conversational speech recognition*, in "arXiv preprint arXiv:1610.05256", 2016