



## Master Sciences Cognitives – 2017-2018

### Développement d'un agent communicant virtuel capable d'apprendre à reconnaître et à produire des séquences de parole à partir d'exemples

La question des relations entre mécanismes de perception et d'action s'est installée au cœur des réflexions et des travaux à la fois en sciences cognitives, avec la découverte des neurones miroirs et les avancées en neuroimagerie, et dans le domaine de la robotique autour des enjeux d'apprentissage et d'adaptabilité.

Nous étudions cette question dans le cadre de la communication parlée, en développant des modèles computationnels d'agents capables à la fois de percevoir et de produire des stimuli de parole simples (voyelles, consonnes, syllabes) et en analysant sur ces modèles en quoi les capacités de production peuvent interagir avec les capacités perceptives, les contraindre, les enrichir ou les compléter.

Ceci a conduit à la mise au point d'un modèle bayésien d'agent communicant, que nous avons baptisé COSMO, pour *Communicating about Objects using SensoriMotor Operations*. Ce modèle probabiliste provient de la modélisation de la situation de communication, dans laquelle deux agents veulent communiquer à propos d'un objet de l'environnement. Le modèle COSMO est basé sur l'hypothèse fondamentale que cette situation de communication peut être internalisée et émulée dans le cerveau de chaque agent, qui est alors en mesure d'agir aussi bien en tant que locuteur qu'auditeur.

Dans ce contexte, l'apprentissage de ce modèle combine l'apprentissage des capacités perceptives et motrices. Il repose sur plusieurs étapes. Dans la première, l'agent apprenant construit un système de catégorisation des signaux acoustiques produits par les agents maîtres (apprentissage auditif). Dans la seconde, l'agent apprend, par accommodation, un système de prédiction des conséquences acoustiques de ses gestes moteurs (apprentissage sensorimoteur).

Enfin, dans la troisième étape, l'agent apprend un système de catégorisation des gestes moteurs lui permettant d'atteindre les cibles acoustiques (apprentissage moteur). Cette séquence d'apprentissages est consistante avec, et inspirée par les données de la littérature sur l'apprentissage de la parole chez le bébé.

#### **Objectifs**

Une grande avancée récente de nos travaux a consisté à montrer que si l'on présentait à un agent COSMO des sons de parole très simples (par exemple des voyelles), il était capable d'inférer après apprentissage les stratégies motrices pour reproduire ce son, et qu'en utilisant ce processus d'inférence, il pouvait mieux traiter les sons bruités qu'un modèle qui ne serait doté que d'un système de traitement direct des sons. **En résumé, on comprend mieux si l'on dispose non seulement d'une oreille mais aussi d'une bouche.** Ce résultat nouveau a été obtenu sur des stimuli très simples et l'objectif est ici de **passer en vraie grandeur**, en mettant en place un système COSMO plus complexe et puissant capable de traiter des séquences de sons (par exemple des séquences de syllabes ou des mots). Ce système devra disposer d'un générateur de sons, un synthétiseur disponible au laboratoire, qui servira de « modèle articulatoire ». L'objectif du stage sera d'implémenter une première version de ce modèle COSMO généralisé à des séquences temporelles, utilisant ce synthétiseur, et de **mettre en place des simulations d'apprentissage dans lesquelles l'agent entend des séquences de parole, apprend à les reproduire, puis, en traitant de nouvelles séquences dégradées, améliore ses performances grâce au processus de reproduction interne.**



### ***Travail proposé***

Pour réaliser l'objectif, l'étudiant devra prendre en charge le synthétiseur articulatoire, l'intégrer dans COSMO, en développant une implémentation de COSMO permettant de traiter des séquences temporelles. Puis il programmera un ensemble de simulations informatiques pour tester les capacités d'apprentissage de ce système, et les potentialités de traitement de stimuli bruités.

### ***Contexte pratique***

Le travail sera réalisé au sein d'une **équipe pluridisciplinaire alliant modélisateurs** – spécialistes de la programmation bayésienne – et **spécialistes de la parole et de la cognition**. Il sera encadré conjointement par Jean-Luc Schwartz, chercheur à GIPSA-Lab (parole) et Julien Diard, chercheur au LPNC (modélisation bayésienne).

**Une poursuite en thèse est envisagée**, dans l'objectif de développer un système complet d'apprentissage sensori-moteur et de traitement de séquences de parole, avec à la fois des **enjeux théoriques** sur le rôle du système moteur dans la perception de la parole, et des **enjeux technologiques** pour la reconnaissance en milieu difficile et le développement d'agents communicants autonomes, dotés de capacités d'apprentissage et d'interaction vocale.

### ***Compétences demandées***

Le candidat devra avoir soit une formation en sciences cognitives et un intérêt pour la modélisation, soit une formation en programmation et simulation informatique, et une affinité pour la modélisation mathématique en sciences cognitives.

### ***Contacts***

Jean-Luc Schwartz (GIPSA-Lab)  
[jean-luc.schwartz@gipsa-lab.grenoble-inp.fr](mailto:jean-luc.schwartz@gipsa-lab.grenoble-inp.fr) 04 76 57 47 12

Julien Diard (LPNC)  
[julien.diard@univ-grenoble-alpes.fr](mailto:julien.diard@univ-grenoble-alpes.fr) 04 76 82 58 93  
<http://diard.wordpress.com>

### ***Financement***

Indemnités de stage assurées par le projet ERC « Speech Unit(e)s »