



Master 2 Sciences Cognitives – 2018-2019

Etude et simulation informatique d'un modèle de perception et production de parole continue Aspects temporels

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 miroir 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, nous avons développé plusieurs instances du modèle COSMO, traitant de questions comme l'évolution des systèmes phonologiques, de la perception de parole dans le bruit, des propriétés fonctionnelles complémentaires de la voie auditive et de la voie motrice, de l'acquisition de répertoires syllabiques et phonémiques, ou encore de l'apparition d'idiosyncrasies dans le développement.

Cependant, **un point commun de ces modèles est leur traitement simplifié de la dimension temporelle**. En effet, les séquences étaient largement restreintes (par ex, une syllabe CV), et leur représentation réduite à des points caractéristiques des trajectoires de parole (par ex, un instant de temps pour la fermeture consonantique et un autre pour le geste vocalique). Cette approche contraste avec la réalité du terrain de la parole naturelle, qui est évidemment constituée de trajectoires acoustiques et articulatoires complexes. Les substrats neuronaux traitant de la parole, dans le cerveau, sont d'ailleurs bien et largement décrits par des modèles à base d'oscillateur neuronal, à même de capturer les propriétés temporelles fines du traitement de la parole et production et en perception.

Objectifs

Notre objectif principal est donc **d'étendre l'architecture du modèle COSMO pour y représenter les aspects temporels**. Pour cela, nous adapterons des outils de modélisation développés dans un autre contexte.

En effet, dans le modèle BRAID de reconnaissance visuelle de mots, nous avons décrit par des chaînes de Markov couplées une architecture hiérarchique allant du stimulus visuel, à la lettre perçue, et à l'identité du mot. En présence d'un stimulus visuel, les distributions de probabilité sur les lettres et les mots perçus évoluent graduellement dans le temps, décrivant une dynamique particulière d'accumulation et de maintien d'information perceptive. Ces distributions évoluent par échange réciproque d'information, rendant compte d'effets bottom-up de propagation d'information sensorielle, combinés à des effets top-down de prédiction d'information contextuelle.

L'objectif est donc une version du modèle COSMO permettant **d'étudier le transfert d'information entre l'entrée sensorielle, une représentation d'éléments distinctifs de la parole (par ex, phonèmes) et une représentation des mots** (ou des séquences de syllabes). Nous étudierons notamment les dynamiques de ces transferts d'information, et les modèles de mélange des flux bottom-up et top-down. Une extension de cette architecture pour y représenter également la voie motrice (c'est une originalité du cadre COSMO) sera



envisagée.

Travail proposé

Pour réaliser l'objectif, l'étudiant devra **implémenter un modèle probabiliste, reprenant et combinant des éléments existants des modèles COSMO et BRAID**. On utilisera ensuite ce modèle dans **des simulations illustratives, permettant de caractériser les propriétés mathématiques du modèle, afin de le comparer aux connaissances expérimentales du domaine**. Une revue de la littérature des modèles computationnels de reconnaissance orale des mots et de la prédiction contextuelle des séquences auditives sera attendue.

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 d'étudier des modèles COSMO plus complets, impliquant des **structures temporelles emboîtées** (syllabes, mots, phrases), **des représentations multimodales** (auditives, somatosensorielles, motrices, voire, visuelles), et permettant de simuler des tâches de perception et de production en ligne, mais aussi **l'apprentissage des répertoires syllabiques et lexicaux** sur la base d'exemples, dans un contexte d'interaction avec un agent maître. Cette thèse comportera des **enjeux théoriques** sur les processus d'interaction et d'apprentissage et leurs corrélats neuro-anatomiques, mais aussi des **enjeux technologiques** pour 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 04 76 57 47 12

Julien Diard (LPNC)
julien.diard@univ-grenoble-alpes.fr 04 76 82 78 07
<http://diard.wordpress.com>

Financement

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