



Sujet de M2 « Sciences Cognitives »

Comparaison de modèles de perception de parole et simulation d'effets comportementaux

Cadre théorique

Le traitement neurocognitif de la parole est classiquement conçu comme une hiérarchie de calculs – comprenant généralement l'extraction de caractéristiques acoustiques ou multisensorielles, la catégorisation pré-lexicale, l'accès lexical, l'intégration prosodique et syntaxique, jusqu'aux étapes finales de compréhension. Il existe dans la littérature de nombreux modèles de perception de la parole et de reconnaissance de mots parlés, souvent inspirés du modèle fondateur Interactive-Activation (IA ; McClelland & Rumelhart, 1981), comme TRACE (McClelland & Elman, 1986) ou PARSER (Perruchet & Vinter, 1998).

Dans le groupe de travail, nous avons développé un modèle probabiliste de perception de la parole (COSMO-Onset ; Nabé et al., 2021). Ce modèle a été défini dans le contexte de la modélisation des mécanismes de détection d'événements temporels du signal de parole, notamment pour le traitement des unités syllabiques. Cependant, ce modèle a une architecture, et repose sur des mathématiques fortement inspirées du modèle de reconnaissance visuelle de mots BRAID (Bayesian model of Word Recognition with Attention, Interference and Dynamics ; Ginestet et al., 2019), lui-même inspiré du modèle IA. Ainsi, une partie du modèle COSMO-Onset a une structure similaire au modèle TRACE : ces deux modèles partagent leur structure hiérarchique, avec des processus feed-forward et feedback entre plusieurs niveaux de représentation. Cependant COSMO-Onset se distingue de TRACE en intégrant un traitement temporel plausible sur le plan neurobiologique, inspiré des avancées récentes dans le domaine des oscillations neuronales et de leur rôle potentiel dans la segmentation du signal acoustique pour la perception de la parole (Poeppel & Assaneo, 2020).

Objectif et Travail proposé

L'objectif principal de ce stage est d'étudier les points communs et différences des modèles COSMO-Onset et des modèles existants inspirés du modèle TRACE, à la fois sur le plan théorique et sur le plan expérimental. Sur le plan théorique, nous comparerons les architectures (espaces de représentation, schémas de propagation de l'information, prise en compte de l'aspect dynamique du stimulus et des traitements), mais aussi les formalismes (modèles connexionnistes vs probabilistes). Par exemple, on pourra « extraire » du modèle COSMO-Onset une « réécriture probabiliste » de TRACE, et montrer l'équivalence.

Sur le plan expérimental, nous chercherons à définir et mettre en œuvre des simulations afin de vérifier la capacité du modèle COSMO-Onset à rendre compte des effets comportementaux classiques en reconnaissance de la parole (par exemple, les effets de fréquence, de voisinage, de lexique, de priming, etc.). Nous nous intéresserons particulièrement aux processus de segmentation dans les processus



psycholinguistiques, et au rôle spécifique que pourrait y jouer la syllabe, qui joue un rôle clé dans les modèles inspirés des oscillations neuronales (Hovsepyan et al., 2020; Ghitza, 2020), alors que l'analyse est centrée sur le niveau inférieur du phone ou du phonème dans le modèle TRACE par exemple.

D'un point de vue pratique, le travail demandé consiste en une revue ciblée de la littérature sur les modèles classiques et les effets à simuler, la prise en main du modèle COSMO-Onset (à la fois conceptuellement, mathématiquement – modélisation probabiliste, écriture d'une variante TRACE-probabiliste – et informatiquement – programmation en Python de cette variante), la conception et l'implémentation informatique d'une ou plusieurs tâches à simuler, et enfin, l'analyse critique et comparative des données simulées.

Compétences demandées

Le candidat devra idéalement avoir une formation en programmation et simulation informatique, et une affinité pour la modélisation mathématique et la psychologie expérimentale : étudiants en Sciences Cognitives ou en Informatique (pour les étudiants d'autres profils, n'hésitez pas à nous contacter). Des connaissances préalables en probabilités sont un plus, mais ne sont pas indispensables.

Contacts

- Julien Diard (LPNC, CNRS) : julien.diard@univ-grenoble-alpes.fr
- Mamady Nabé (LPNC, doctorant) : mamady.nabe@univ-grenoble-alpes.fr
- Jean-Luc Schwartz (GIPSA-Lab, CNRS) : jean-luc.schwartz@gipsa-lab.grenoble-inp.fr

Financement

Une indemnité de stage est assurée.

Références

- Ghitza, O. (2020). "Acoustic-driven oscillators as cortical pacemaker": A commentary on Meyer, Sun & Martin (2019). *Language, Cognition and Neuroscience*, 1–6.
- Ginestet, E., Phénix, T., Diard, J., & Valdois, S. (2019). Modeling the length effect for words in lexical decision: The role of visual attention. *Vision research*, 159, 10–20.
- Hovsepyan, S., Olasagasti, I., & Giraud, A.-L. (2020). Combining predictive coding and neural oscillations enables online syllable recognition in natural speech. *Nature communications*, 11(1), 1–12.
- McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: I. an account of basic findings. *Psychological review*, 88(5), 375.
- McClelland, J. L., & Elman, J. L. (1986). The trace model of speech perception. *Cognitive psychology*, 18(1), 1–86.
- Nabé, M., Schwartz, J.-L., and Diard, J. (2021). COSMO-Onset: a neurally-inspired computation model of spoken word recognition, combining top-down prediction and bottom-up detection of syllabic onsets. *Frontiers in Systems Neuroscience*, 15:653975.
- Perruchet, P., & Vinter, A. (1998). Parser: A model for word segmentation. *Journal of memory and language*, 39(2), 246–263.
- Poeppl, D., & Assaneo, M. F. (2020). Speech rhythms and their neural foundations. *Nature Reviews Neuroscience*, 1–13.