Machine learning models of children's multimodal communication

Supervisors:
Dhia-Elhak Goumro
Leonor Becerra
Abdellah Fourtassi

Aix-Marseille University

An M2 internship position is available within the TALEP team at the Laboratoire d'Informatique & Systèmes (LIS) lab based in Luminy.

The goal of this internship is to use state-of-the-art machine-learning models to study how children develop their ability to communicate in a multimodal way (i.e., using visual cues to learn turn-taking); a crucial question in developmental cognitive science, with important implications for child-computer interaction systems.

We will work on our new corpus (ChiCa) collected in Marseille with children conversing with their parents both face-to-face (using a mobile eye-tracking device) and via Zoom across three age groups in middle childhood: 7, 9, and 11 years old. The corpus was fully transcribed and annotated for eye gaze.
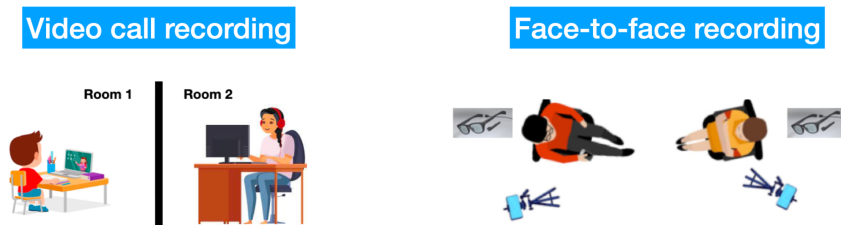
Regarding the models, we will train multimodal deep learning models on the data available from three modalities (audio, visual & verbal) in the corpus. The models will be trained to learn and predict children's behavior given signals from the caregiver. The goal is to identify the role of each modality (and their combination) at different developmental ages. The long-term impact is to improve child-oriented conversational AI in different fields, including education and health.

The internship requires the mastery of Python programming language and a solid background (and ideally experience) in using deep learning models.

If interested, please send an email (together with a "relevé de notes" and/or any other document that you deem useful to appreciate your skills) to Abdellah Fourtassi at (abdellah.fourtassi@gmail.com)

Some References:

[1] Agrawal, A., Liu, J., Bodur, K., Favre, B., & Fourtassi, A. (2023). Development of Multimodal Turn Coordination in Conversations: Evidence for Adult-like behavior in Middle Childhood. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 45, No. 45).

[2] Das, A., Kottur, S., Gupta, K., Singh, A., Yadav, D., Moura, J. M., ... & Batra, D. (2017). Visual dialog. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 326-335).

[3] Zadeh, A., Liang, P. P., Poria, S., Vij, P., Cambria, E., & Morency, L. P. (2018, April). Multi-attention recurrent network for human communication comprehension. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 32, No. 1).