A Multimodal Machine Learning Approach To Study Convergence Phenomena In Natural Conversations

Supervisor : Leonor Becerra
Laboratoire d'Informatique et Systèmes
Aix-Marseille Université

Description of the PhD thesis project

Natural conversations are characterized by *convergence* phenomena between participants: during a conversation, interlocutors become progressively *aligned* at different levels [Pickering and Garrod, 2021]. For example, speakers imitate some characteristics of their interlocutor's speech sound (prosody level), they use identical vocabulary (lexical level), they repeat or paraphrase structures (syntactic level), etc. Mutual understanding and more generally the *success of an interaction* depends on such convergence behaviors.

Moreover, interaction between humans is *multimodal* by nature. When people interact with each other in a conversation, information from different modalities comes into play: verbal (use of language), visual (gestures, gaze, etc.), physiological (heart rate, skin temperature, etc.) and cerebral (neural correlates between the different signals) [Bogels 2020; Pickering and Garod, 2021]. Therefore, in order to have a complete vision of human interaction, it is crucial not to focus only on one part of these data.

The first goal of this PhD project is to gather a large set of convergence features, coming from all *verbal* and *non-verbal modalities*, into a **predictive model of convergence levels** at local and global scales, by using **multimodal machine learning** techniques [Guo et al., 2019; Baltrušaitis et al., 2019]. In a second step, we propose to study the correlation between the convergence level and the coordination of brain activities between the participants (also called **brain-to-brain entrainment**) [Perez Fernández et al., 2017]. The hypothesis is that a high level of convergence could be correlated with a higher coupling at the brain level.

In terms of **methodology**, we will start by analyzing an existing dataset named Badalona-EPSN containing audio, video and cerebral signals recorded in a natural situation and enriched with automatic annotations in the linguistic and gestural domains. The convergence level is manually annotated (work in progress). Moreover, this dataset will be completed soon by data acquired on the context of another project. Feature extraction and modeling will be applied to these two datasets, representing a total of 5h, fully annotated at all levels.

Therefore, we propose an **interdisciplinary** PhD thesis project which involves three main **originalities**: 1) This thesis will address for the first time the development of a predictive model of convergence levels between participants in a natural conversation; 2) The information from the *cerebral signal* is integrated, in addition to the verbal and visual modality, which will bring new insights into the brain processes involved for a successful interaction; 3) The results of this thesis will allow innovative applications in other domains, such as Human-Computer interaction.

Expected profile of the candidate

- Knowledge in at least two of the following areas: Machine Learning, Natural Language Processing and Image/Signal Processing
- Programming skills in Python
- Ability to communicate effectively in English, both orally and in writing
- Curious, autonomous, rigorous mind

Bibliography

T. Baltrušaitis, C. Ahuja and L-P. Morency, “Multimodal Machine Learning: A Survey and Taxonomy”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume 41, Issue 2, 423-443, 2019.

S. Bögels, “Neural correlates of turn-taking in the wild: Response planning starts early in free interviews”, Cognition, vol. 203, 104-347, 2020.

W. Guo, J. Wang and S. Wang, “Deep Multimodal Representation Learning: A Survey”, in IEEE Access, vol. 7, 63373-63394, 2019.

A. Pérez Fernández, M. Carreiras, and J. A. Duñabeitia, “Brain-to-brain entrainment: EEG interbrain synchronization while speaking and listening”, Scientific Reports , vol. 7, 2017.

M. Pickering and S. Garrod, “Understanding Dialogue”. Cambridge University Press, 2021.