

A multimodal approach to study convergence phenomena in natural conversations

Supervisors : Leonor Becerra, Philippe Blache et Eliot Maës
Aix-Marseille Université

Description of the project

During natural conversations, interlocutors become progressively *aligned* at different levels. 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). Therefore, in order to have a complete vision of human interaction, it is crucial not to focus only on one part of these data.

We propose in this project to analyze an existing dataset named Badalona-EPSN containing audio, video and cerebral signals recorded in a natural situation. This corpus is enriched with automatic annotations in the linguistic and gestural domains. The goal of the project is to start exploring alignment/entrainment at different modalities (lexicon, syntax, discourse) and study their possible correlation with the neurophysiological signal. This project will be carried out in collaboration with two researchers and one PhD student, within a larger project on the study of brain traces of language phenomena.

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.

P. Blache, S. Antoine, D. De Jong, L-M. Huttner, E. Kerr¹, T. Legou, E. Maës, C. François, "The Badalona Corpus An Audio, Video and Neuro-Physiological Conversational Dataset", LREC, 1-8, 2022.

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

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