Disentagling Prediction and Production in Neural Tracking via multivariate TRFs in Natural Conversation

Maïwenn Fleig[1,2,3], Shuai Wang[4], Anna E. Dudek[4,5], Jean-Marc Freyermuth[4], Leonor Becerra[1,3], Philippe Blache[2,3]

[1]Aix Marseille University, CNRS, LIS, France, [2]Aix-Marseille University, CNRS, LPL, France, [3]Institute of Language, Communication and the Brain, France, [4]Aix Marseille University, CNRS, I2M, France, [5]AGH University of Krakow, Faculty of Applied Mathematics, Poland

Human language understanding is inherently predictive: during conversation, listeners anticipate upcoming words using context and prior knowledge. While predictive neural mechanisms are well-studied in controlled settings, their role in spontaneous conversation remains unclear since passive paradigms miss the dynamic interplay of comprehension, production, and adaptation. Here, we analyze EEG recordings from interacting speakers during natural dialogue. This setting presents a key methodological challenge: neural signals for language comprehension are deeply intertwined with the activity of producing one's own speech. Natural conversation may also introduce additional movement-related noise (e.g., eye, head, and facial muscle activity), but the central challenge is distinguishing prediction-related responses from concurrent production-related processes without excluding speaking intervals, which would remove behaviorally meaningful events of the interaction.

To address this, we apply multivariate temporal response functions (mTRFs) to disentangle predictive cortical tracking from speech production-related activity. We analyze a natural conversational dataset in which two speakers engage in free turn-taking while their EEG is recorded [1]. We model one interlocutor's (speaker B) EEG as a function of linguistic features derived from the other speaker's speech (speaker A), alongside the word onset capturing the participant's own speech production (see Figure 1). These include word onset, surprisal [3] and semantic dissimilarity [2], capturing contextual (un)predictability and meaning evolution. This modeling approach allows us to isolate predictive linguistic processing in naturalistic interaction while accounting for production-related neural activity. For the present study, we analysed 19 dyads (38 participants, 17.4 min on average) with EEG signals acquired using two 64-channel Biosemi active electrode systems.

We examine the estimated model weights, which relate stimulus fluctuations to EEG activity and quantify coupling between linguistic input and neural responses. We focus specifically on weights associated with linguistic features from speaker A's speech, as these reflect how predictive features relate to speaker B's neural response. TRF significance is assessed using a mass-univariate spatiotemporal cluster-based permutation test [4]. Looking at the word-onset, responses to speaker B's word onsets emerge earlier, even before onset, whereas responses to speaker A's word onsets occur at around 50 ms and are consistent with the timing of the N400 complex (see Figure 2). Surprisal and semantic dissimilarity do not show significant effects, which is consistent with prior findings showing that these higher-level linguistic effects are typically more difficult to detect. However, since natural conversations lack clear sentence structure, which may obscure these effects, future research should explore conversation-level methods for calculating this metric specifically.

Figures

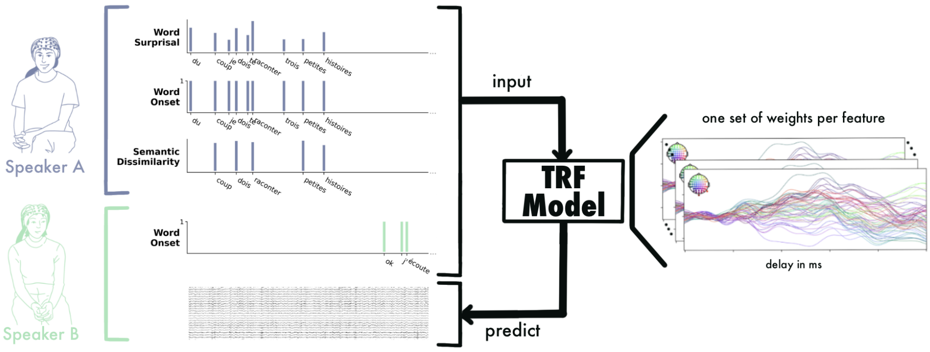

Figure 1: Overview of the experimental framework. Features were extracted from the transcriptions of both speakers and used as model inputs. EEG recordings from Speaker B were predicted using TRF models trained on features from both speakers, implicitly controlling for their mutual influence.

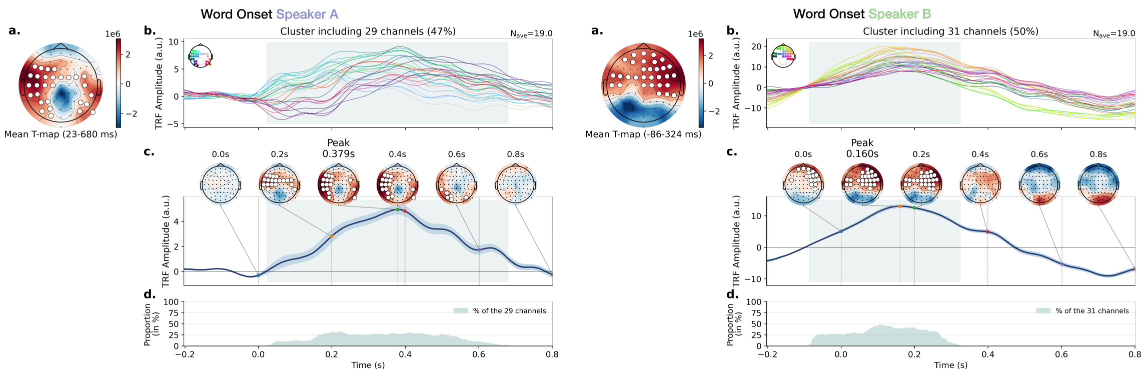

Figure 2: *Left* panel shows responses time-locked to Speaker A's word onsets, reflecting prediction of Speaker B;=. *Right* panel shows Speaker B's own productions, with significant clusters preceding word onset (0.0 s). **a.** Scalp topography of the averaged T-map during the significant window; white circles mark significant electrodes and colors indicate statistical magnitude. **b.** Averaged feature response time course with the significant window shaded. **c.** Averaged mTRF time course from significant-cluster channels with variance; insets show weight topographies at three latency peaks. **d.** Proportion of significant-cluster channels contributing to the spatiotemporal cluster over time.

References

[1] Auriane Boudin, Roxane Bertrand, Stéphane Rauzy, Matthis Houlès, Thierry Legou, Magalie Ochs, and Philippe Blache. Smyle: A new multimodal resource of talk-in-interaction including neuro-physiological signal. In *Companion Publication of the 25th International Conference on Multimodal Interaction*, pages 344–352, 2023.

[2] Michael P Broderick, Andrew J Anderson, Giovanni M Di Liberto, Michael J Crosse, and Edmund C Lalor. Electrophysiological correlates of semantic dissimilarity reflect the comprehension of natural, narrative speech. *Current Biology*, 28(5):803–809, 2018.

[3] John Hale. A probabilistic Earley parser as a psycholinguistic model. In *Second Meeting of the North American Chapter of the Association for Computational Linguistics*, 2001. URL `https://aclanthology.org/N01-1021/`.

[4] Eric Maris and Robert Oostenveld. Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, 164(1):177–190, August 2007. ISSN 01650270.