A novel transformer-based approach for estimating causal interaction in multichannel electroencephalographic data

SAPIENZA
UNIVERSITÀ DI ROMA

Luca Maurici 1809678

Master Thesis – M.Sc. in Artificial Intelligence and Robotics

Relatore: Prof.ssa Laura Astolfi

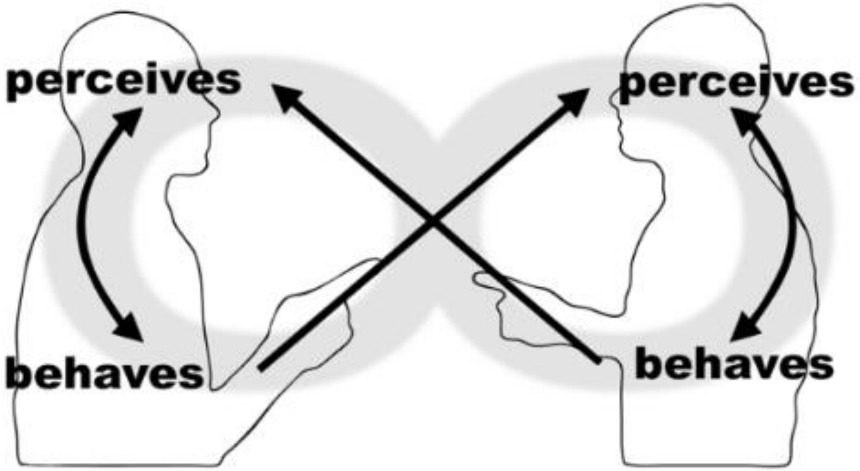
Correlatore: Prof. Nicola Toschi

Controrelatore: Prof. Luca Iocchi

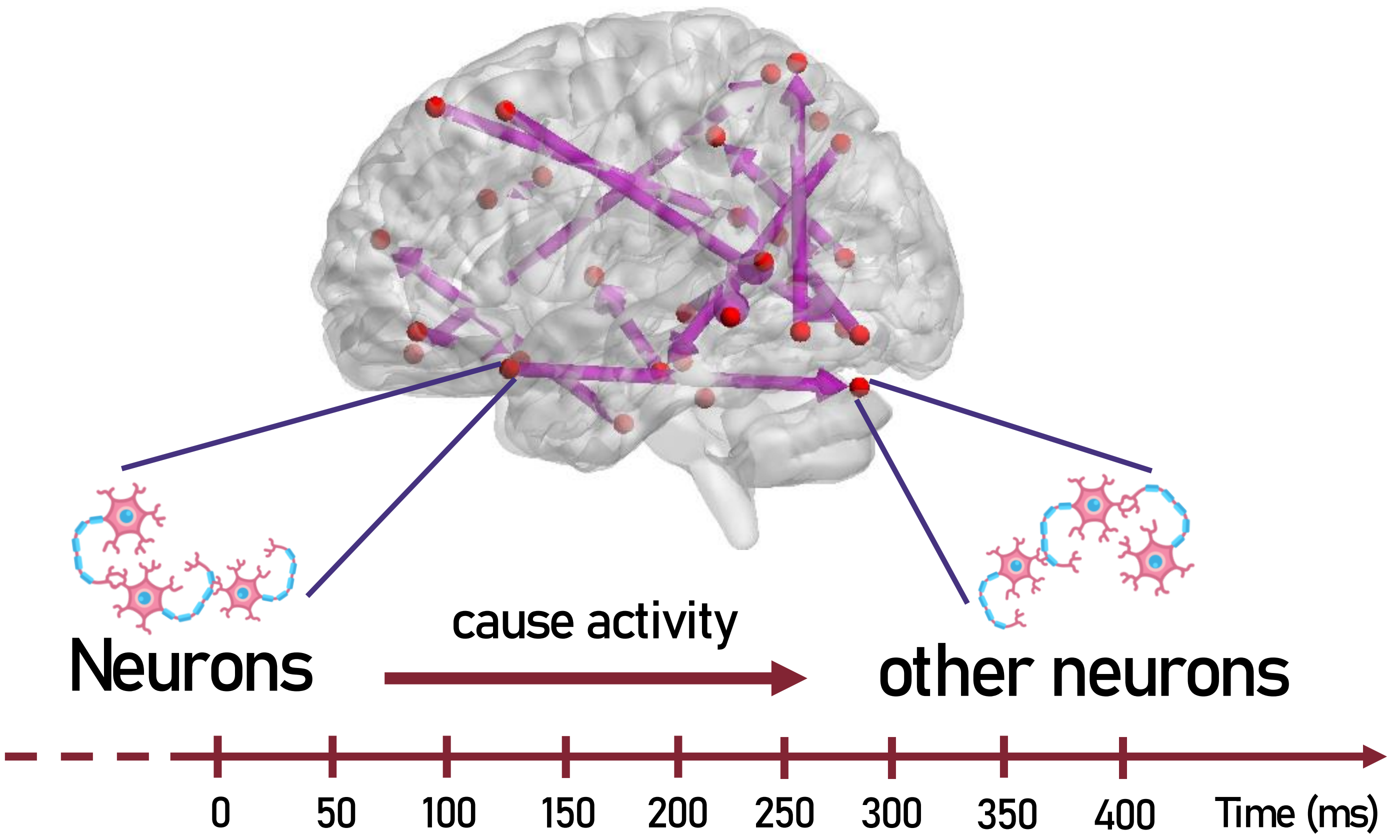
25-10-2022

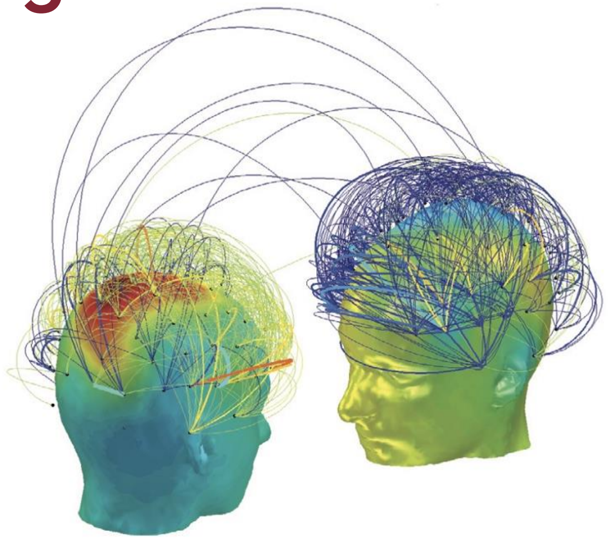
Introduction – EEG Hyperscanning

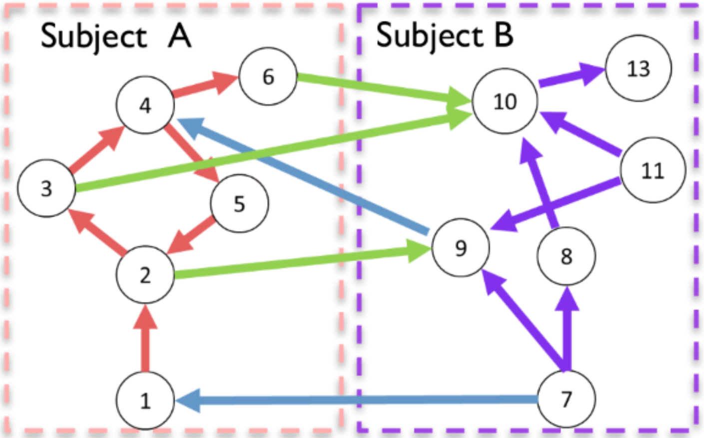
- Simultaneous acquisition of brain signals on two (or more) subjects during their interaction.
- Interdependencies between the two sets of brain signals can reveal brain processes that support the interaction.
- EEG hyperscanning allows more natural, face-to-face interaction and follows the brain dynamics with an excellent time resolution.



Introduction – Causality in hyperscanning

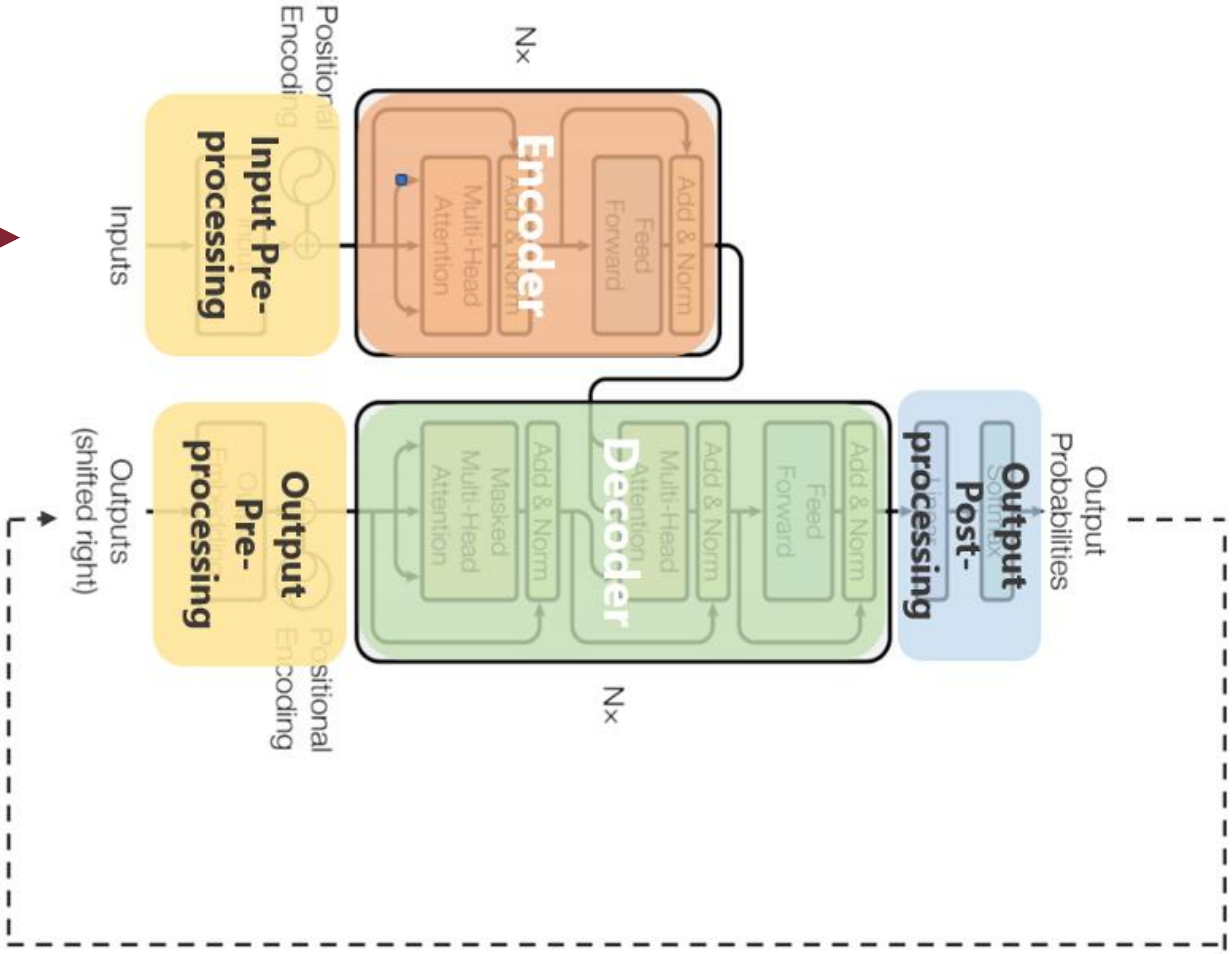






Introduction – Basic of Transformers

The **pen** is
on the **table**



La penna è
sul tavolo

Aims

Propose two novel methods

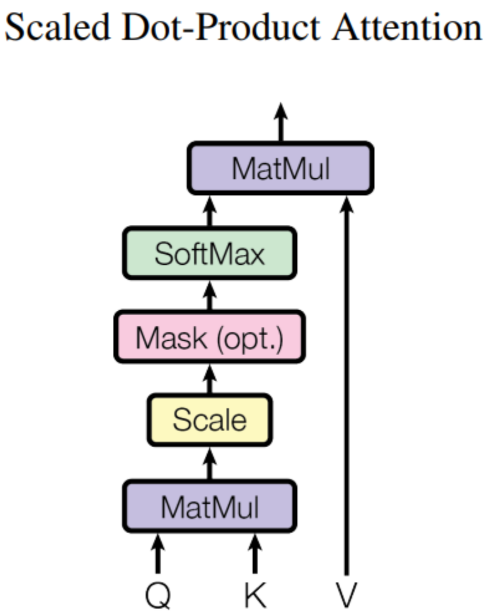
Method 1: Conditioned Granger Causality based on Spacetimeformer residuals

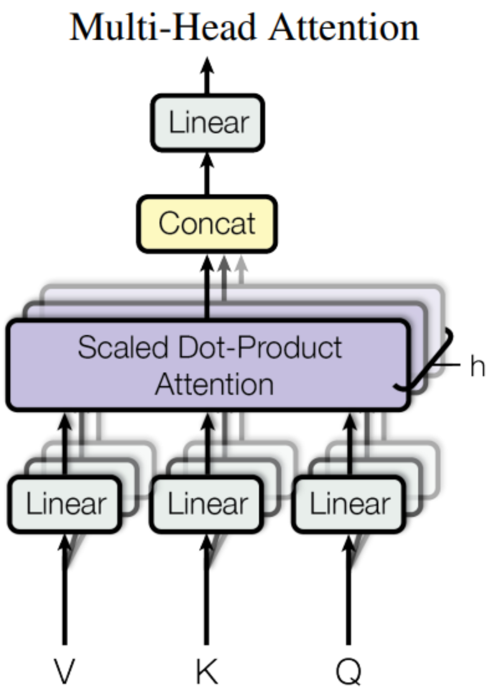
- Assess feasibility of the **novel approach**, using synthetic EEG data as a benchmark, to compare results with a ground truth.

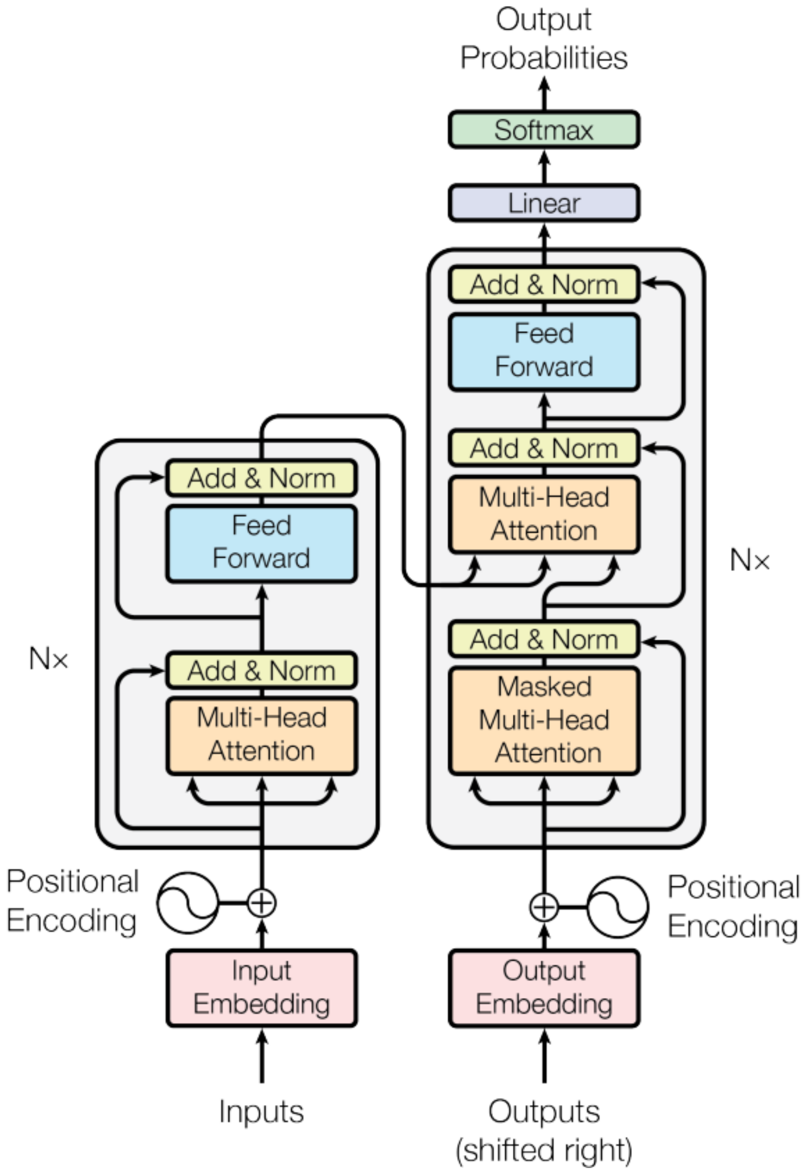
Method 2: Attention matrices as measure of causality

- Assess the plausibility of this **novel method** by applying it to **real** hyperscanning EEG data.

Methods – Transformers

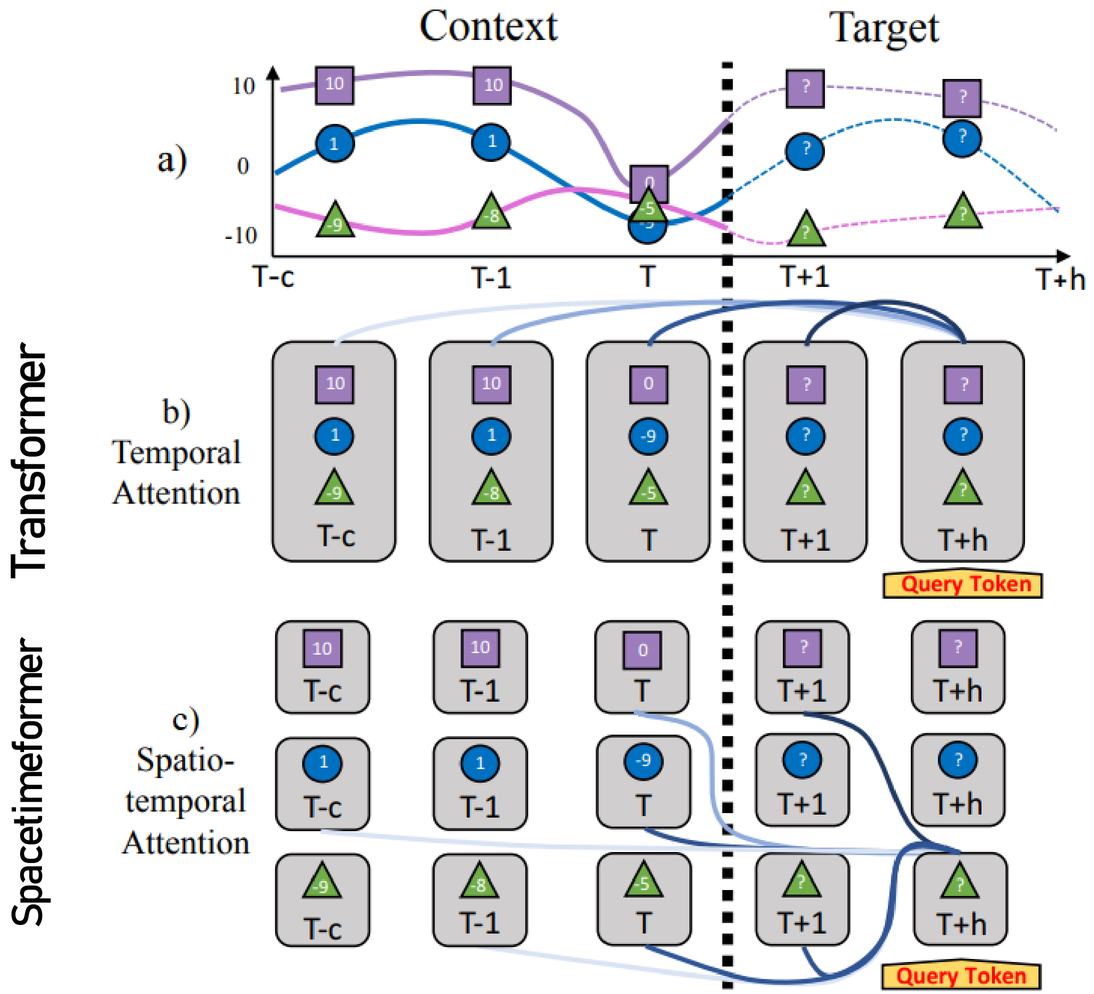


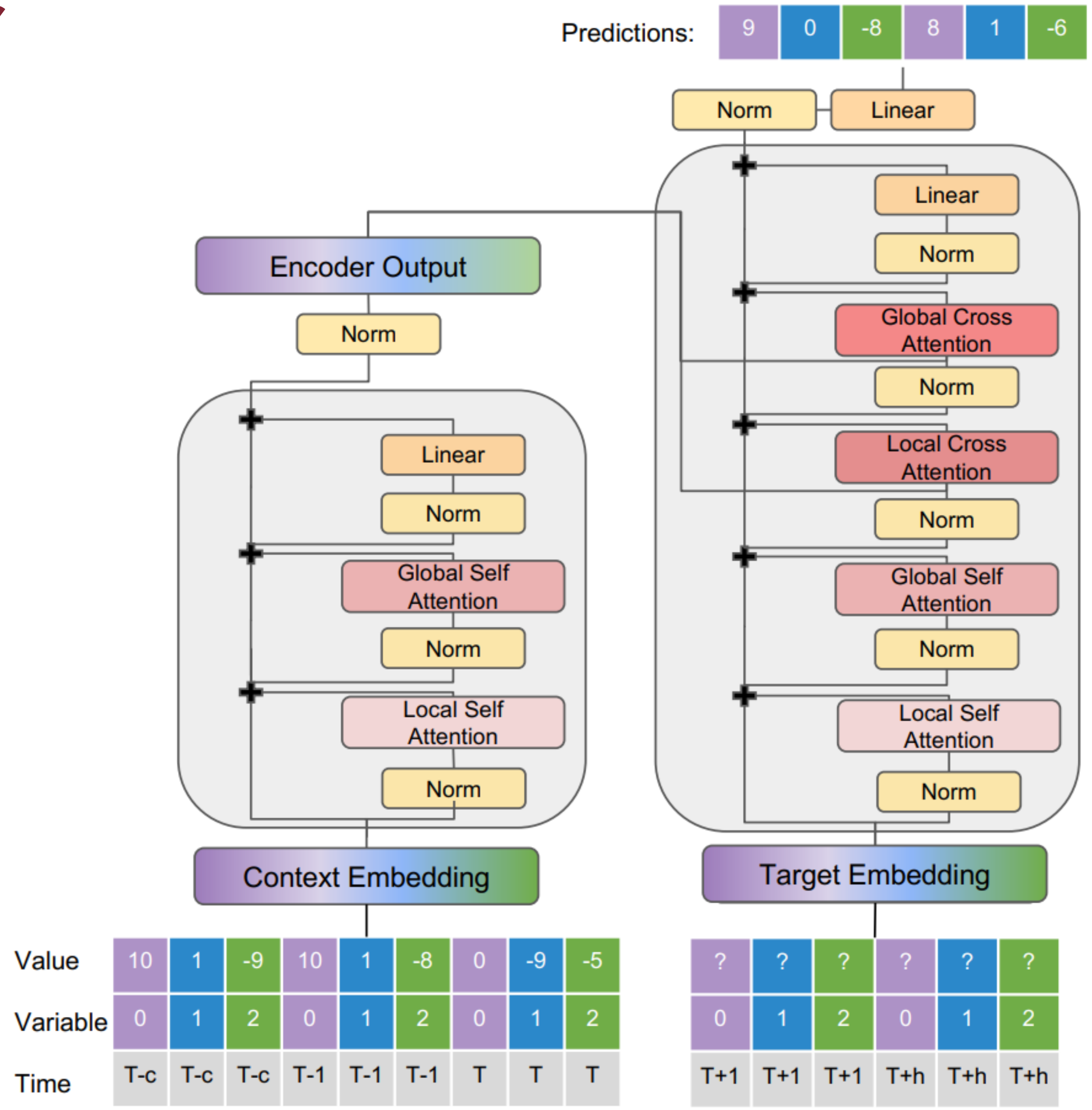




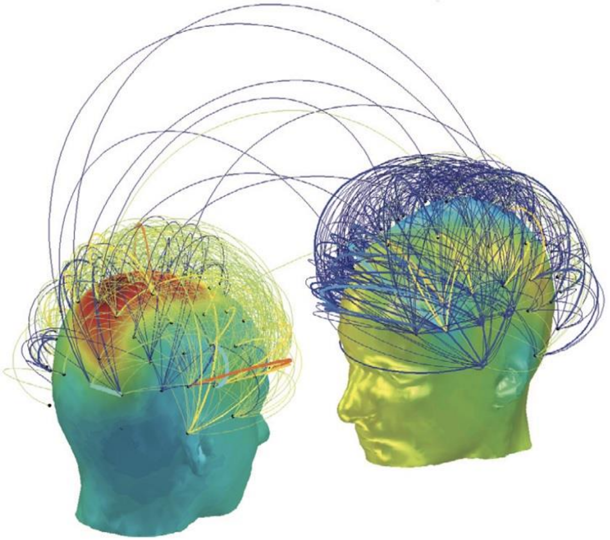
$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

Methods - Spacetimeformer

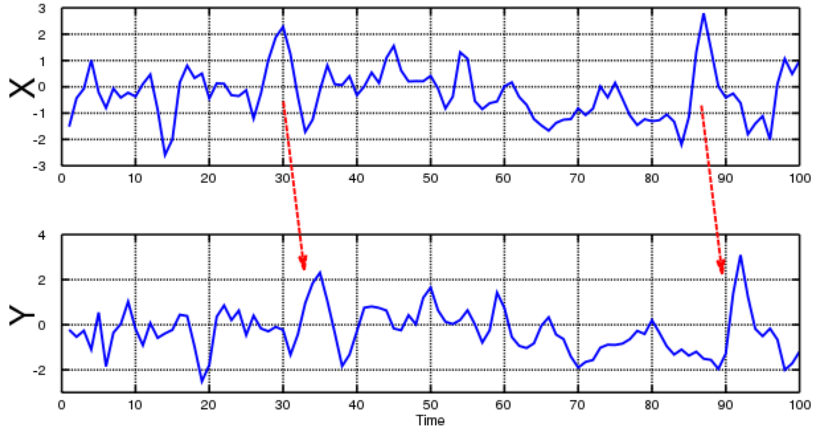




Methods – Wiener-Granger causality



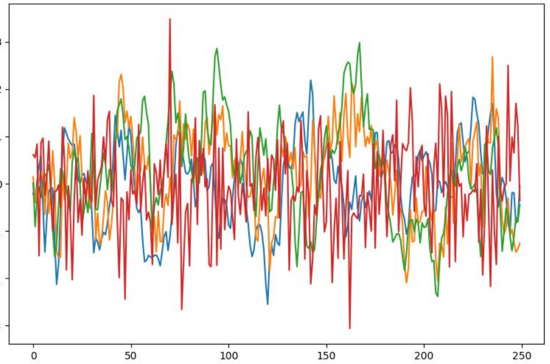
Granger causality is based on the improvement of the predicting capability of a model by incorporating the past of a second signal. This is measured by the reduction of the residuals.

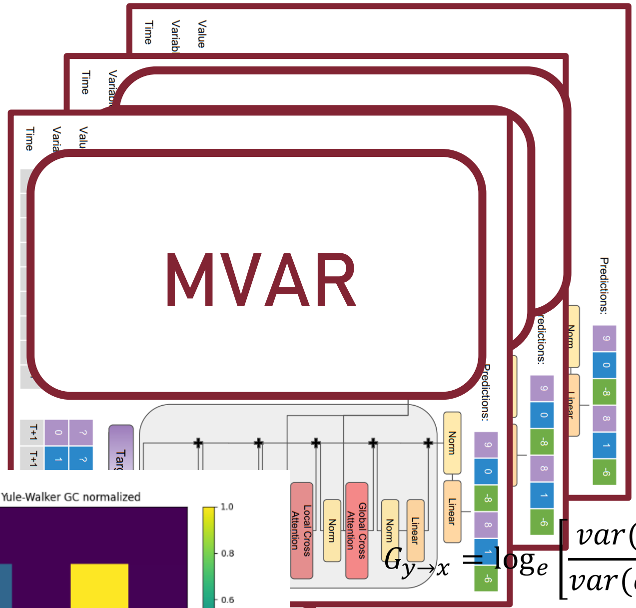


$$G_{y \to x} = \log_e \left[\frac{var(e_x)}{var(e_{yx})} \right]$$

Method 1: CGC with Spacetimeformer

Synthetic EEG
generation

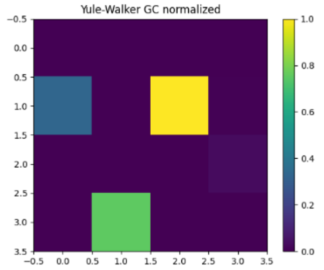




Forecasted EEG

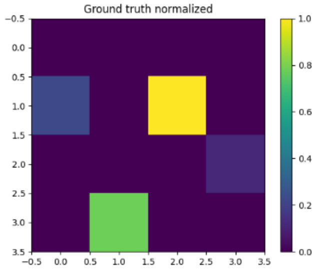
$$G_{y \to x} = \log_e \left[\frac{var(e_x)}{var(e_{yx})} \right]$$

$$\mathrm{Var}(E(n))$$



Causality matrix

COMPARISON



Ground truth

Results of Method 1

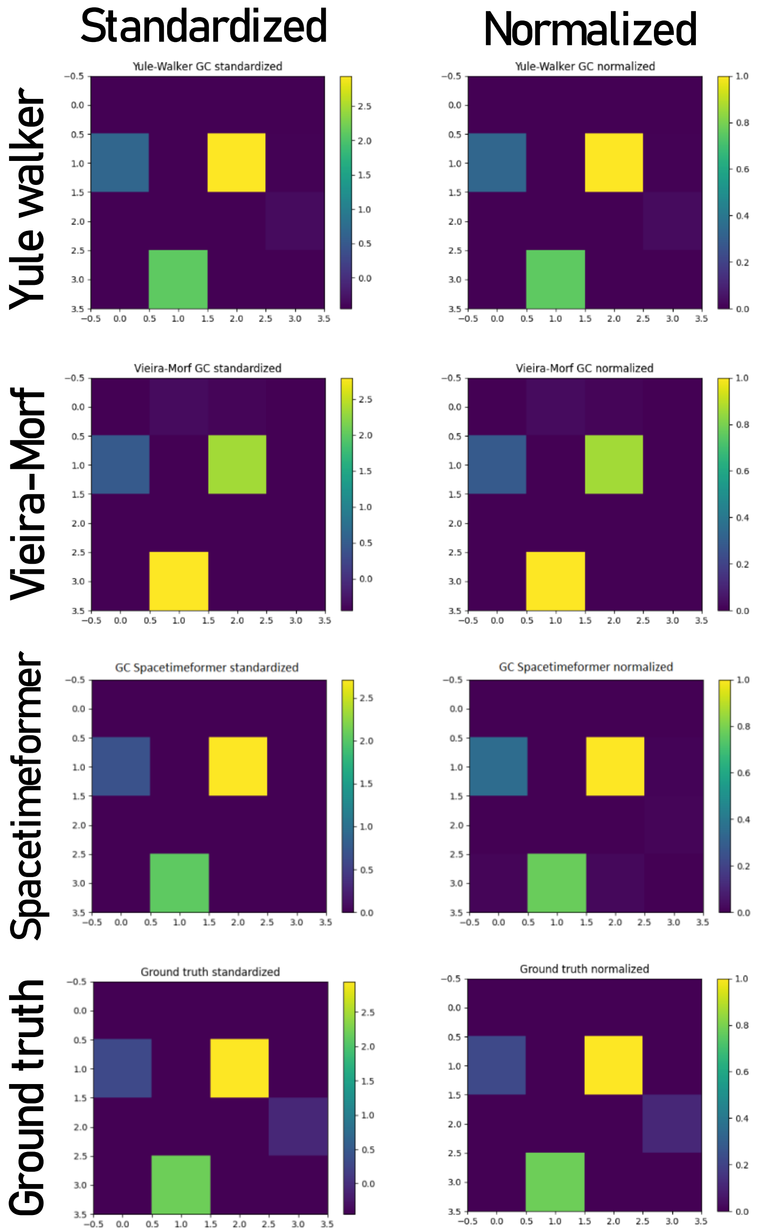
	Negative to zero thresholding after standardization			Negative to zero thresholding before normalization		
	Yule-Walker	**Vieira-Morf**	**Spacetimeformer**	**Yule-Walker**	**Vieira-Morf**	**Spacetimeformer**
	0.35	0.35	0.35	0.29	0.30	0.32
	0.59	0.57	0.51	0.49	0.48	0.48
	0.25	0.26	0.23	0.23	0.24	0.23
	0.42	0.30	0.37	0.39	0.30	0.37
	0.08	0.22	0.14	0.10	0.24	0.12
	0.59	0.57	0.56	0.49	0.48	0.51
	0.25	0.26	0.26	0.23	0.24	0.27
	0.42	0.30	0.38	0.39	0.30	0.39
	0.08	0.22	0.13	0.10	0.24	0.10
	0.22	0.24	0.14	0.19	0.29	0.20
MEAN	**0.323**	**0.330**	**0.308**	**0.291**	**0.312**	**0.298**
STD. DEV	**0.18**	**0.13**	**0.15**	**0.15**	**0.09**	**0.14**

p = 0.160

p = 0.153

p = 0.167

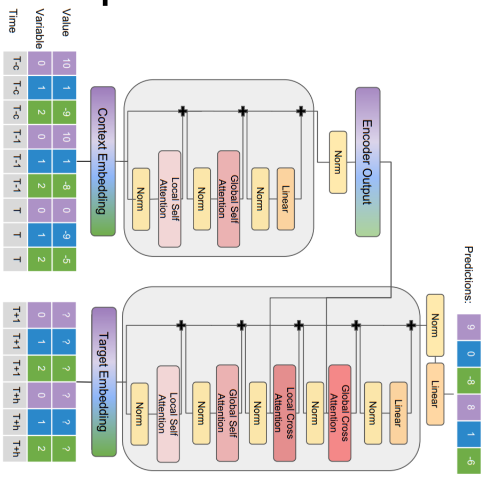
p = 0.288



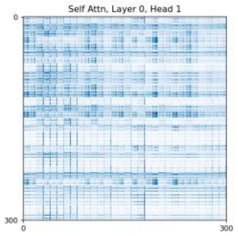
Method 2: Attention matrices as causality measure

Input EEG

Spacetimeformer

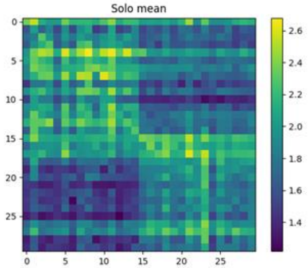


Forecasted EEG



Attention matrix

aggregations

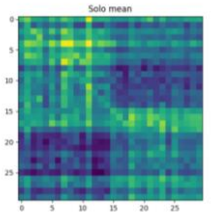


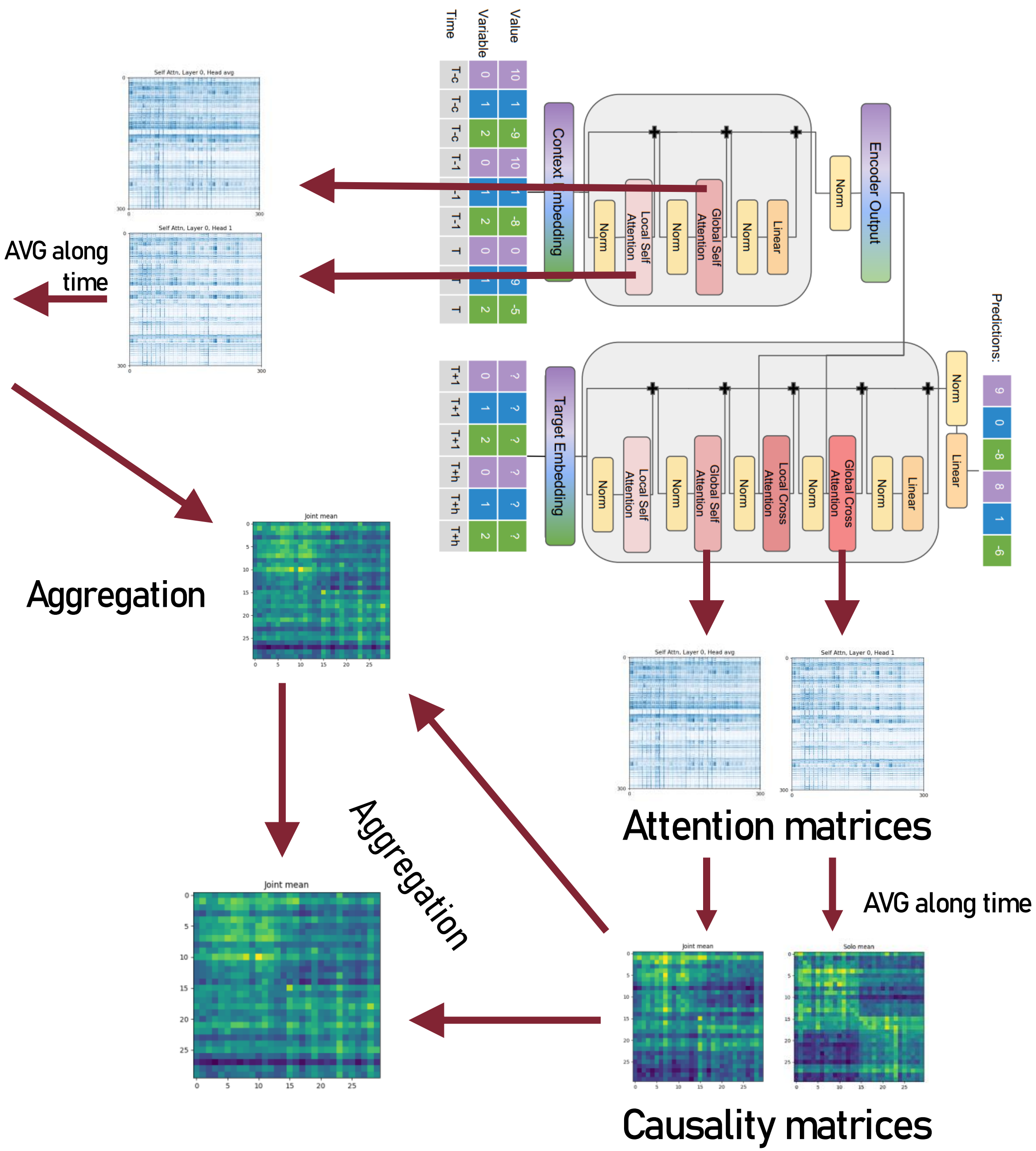
Causality matrix

Method 2: Attention matrices

Types of aggregation considered

1. Cross-attention on Layer 0
2. Cross-attention on Layer 1
3. Average of cross-attentions
4. Element-wise product between cross-attentions
5. Self-attention on Layer 0
6. Self-attention on Layer 1
7. Average of self-attentions.
8. Element-wise product between self-attentions
9. Average of all the attention on every layer
10. Element-wise product between all the attention on every layer
11. Element-wise product between the averages of cross and self-attention





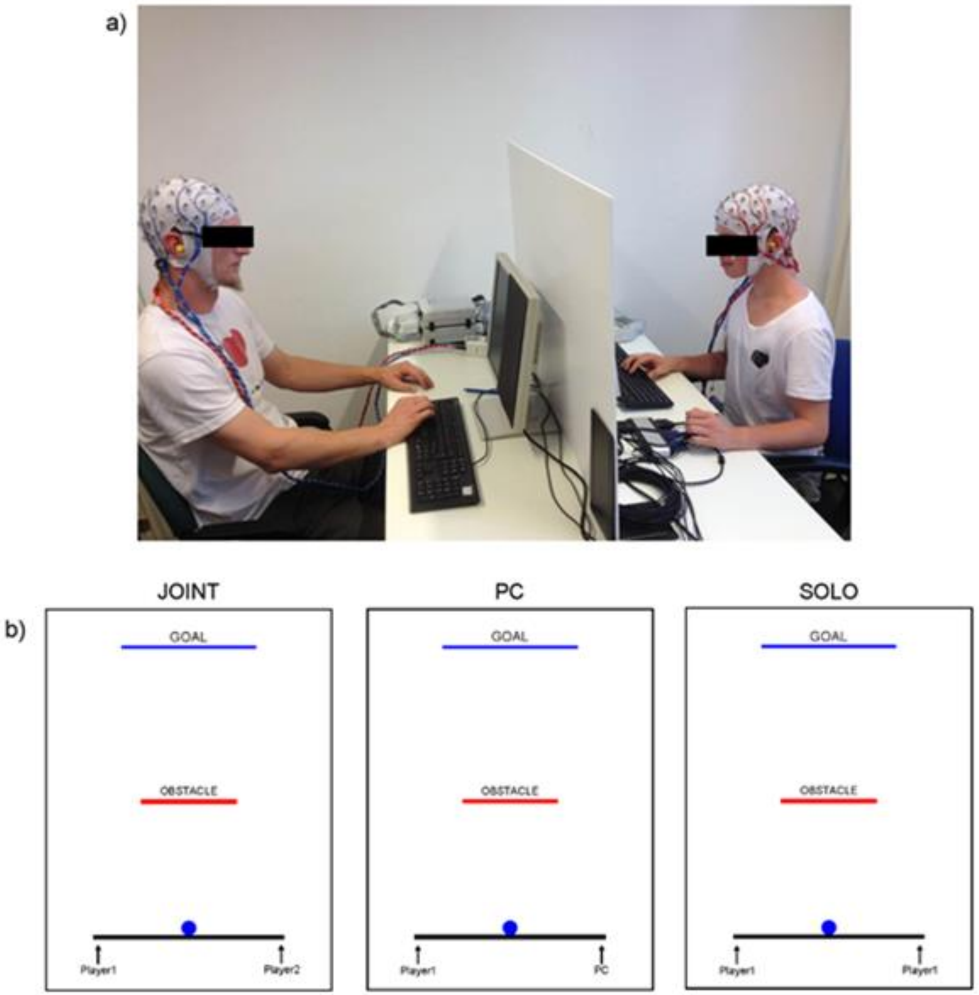
Method 2: Hyperscanning Joint experiment

Two conditions considered:

1. **Solo** experiment: **each subject** completed the task on **their own**, two fingers to **control both sides** of the virtual **bar**.
2. **Joint** experiment: the dyad **worked** on the same task **together**. **Each** participant **used one finger** to press a button to **control one side** of the virtual **bar**.

(Astolfi, et al., 2020)

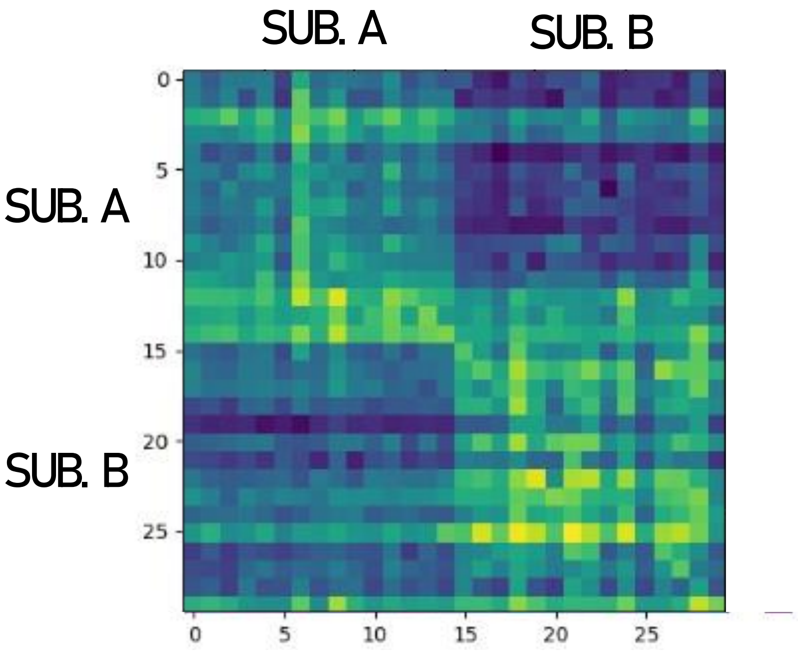
The neuroelectrical hyperscanning recordings were performed with a 128-channel EEG acquisition system (64+64 channels) with a sampling frequency of 250 Hz



Method 2: Graph's indices

Six graph's indices

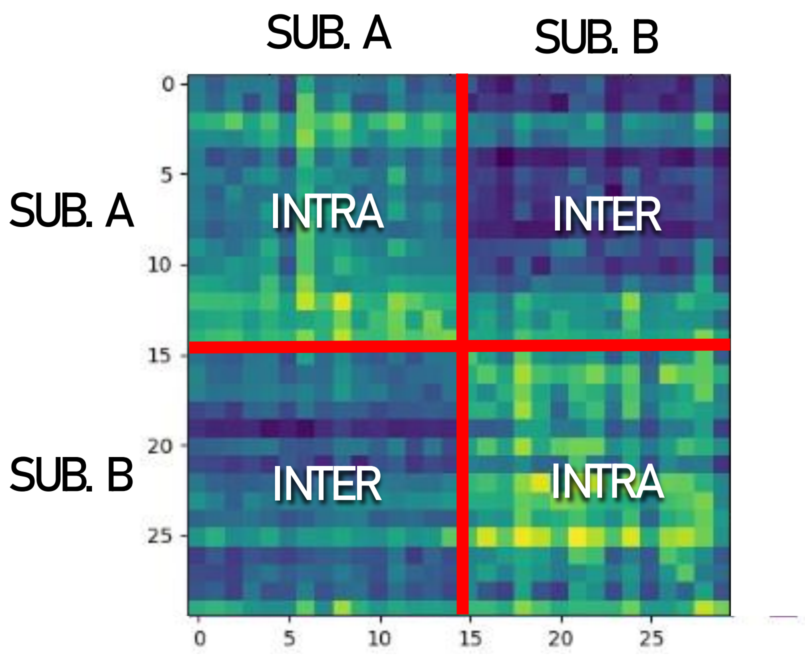
1. **Sum** of the **intra**-group connections
2. **Sum** of the **inter**-group connections
3. Weighted density of the **intra**-group connections
4. Weighted density of the **inter**-group connections
5. Divisibility
6. Modularity



Method 2: Graph's indices

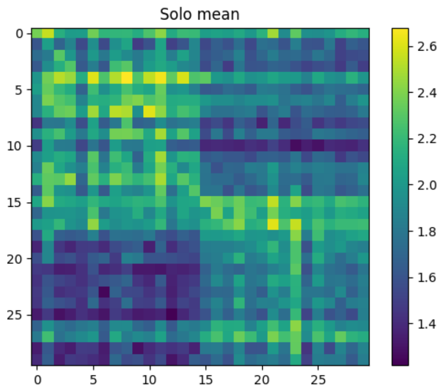
Six graph's indices

1. **Sum** of the **intra**-group connections
2. **Sum** of the **inter**-group connections
3. Weighted density of the **intra**-group connections
4. Weighted density of the **inter**-group connections
5. Divisibility
6. Modularity



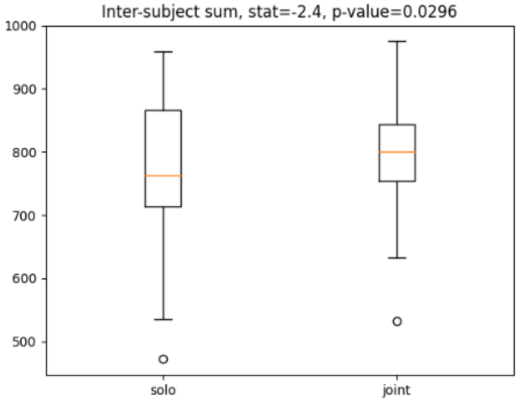
Results of Method 2 – Table of t-tests

Matrix type	Sum intra-sub.		Sum inter-sub.		Density intra-sub.		Density inter-sub.		Divisibility		Modularity	
	Stat.	P-value	Stat.	P-value	Stat.	P-value	Stat.	P-value	Stat.	P-value	Stat.	P-value
1. Cross-attention on Layer 0	1.69	0.1119	-1.69	0.1119	1.47	0.1613	-1.47	0.1613	1.47	0.1613	0.34	0.7381
2. Cross-attention on Layer 1	-0.99	0.3366	0.99	0.3366	-0.90	0.3828	0.90	0.3828	-0.85	0.4089	-0.91	0.3752
3. Average of cross-attentions	1.14	0.2704	-1.14	0.2704	1.24	0.2340	-1.24	0.2340	1.27	0.2251	0.22	0.8262
4. Element-wise product between cross-attentions	-0.27	0.7902	0.27	0.7902	-0.35	0.7348	0.35	0.7348	-0.30	0.7705	-0.42	0.6771
5. Self-attention on Layer 0	1.05	0.3121	-1.05	0.3121	1.28	0.2191	-1.28	0.2191	1.30	0.2143	-0.47	0.6433
6. Self -attention on Layer 1	2.40	**0.0296**	-2.40	**0.0296**	2.59	**0.0204**	-2.59	**0.0204**	2.62	**0.0192**	1.64	0.1224
7. Average of self-attentions	2.04	0.0592	-2.04	0.0592	2.28	**0.0377**	-2.28	**0.0377**	2.31	**0.0355**	1.08	0.2953
8. Element-wise product between self-attentions	2.29	**0.0369**	-2.29	**0.0369**	2.61	**0.0197**	-2.61	**0.0197**	2.66	**0.0180**	1.28	0.2209
9. Average of all the attention on every layer	1.78	0.0950	-1.78	0.0950	1.82	0.0894	-1.82	0.0894	1.84	0.0854	0.49	0.6298
10. Element-wise product between all the attention on every layer	0.34	0.7364	-0.34	0.7364	0.15	0.8810	-0.15	0.8810	0.34	0.7356	-0.69	0.5007
11. Element-wise product between the averages of cross and self-attention	1.93	0.0725	-1.93	0.0725	1.91	0.0757	-1.91	0.0757	1.95	0.0702	0.14	0.8928

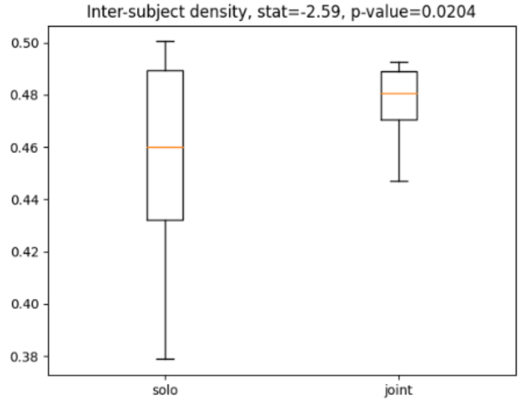


Results of Method 2 – Self-attention on Layer 1

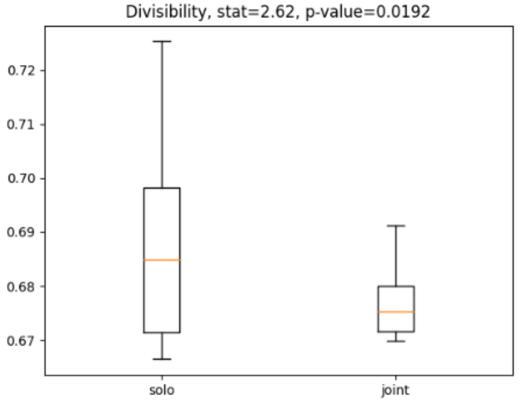
Inter-subject sum



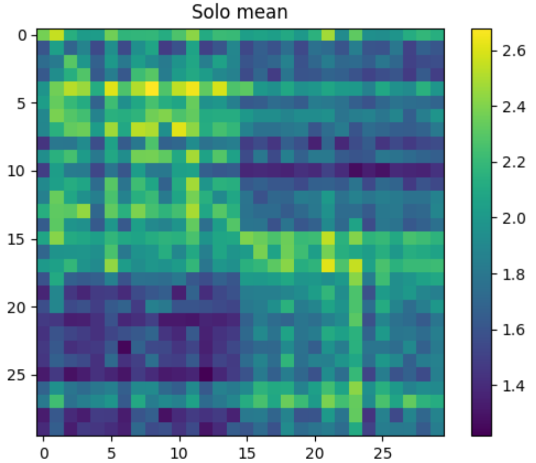
Inter-subject density



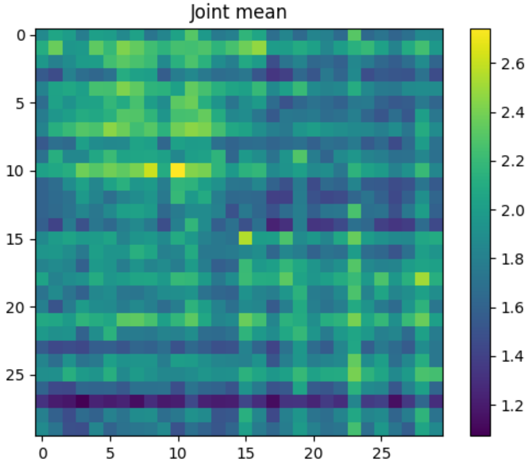
Divisibility



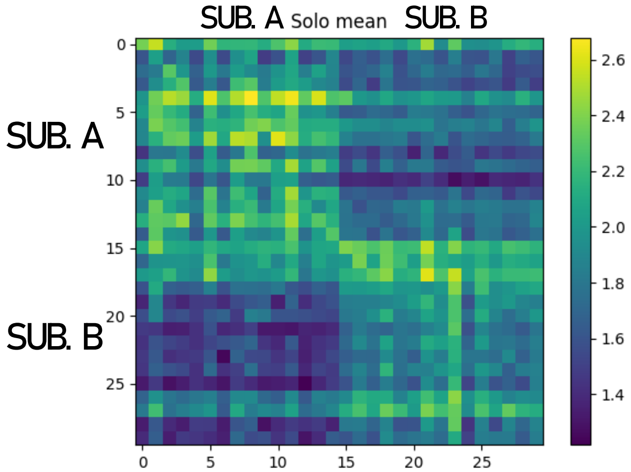
Solo condition



Joint condition

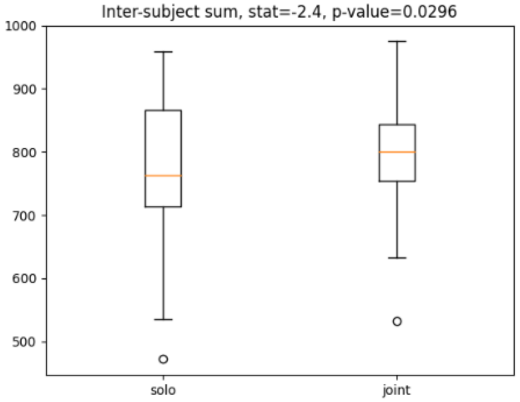


SOLO

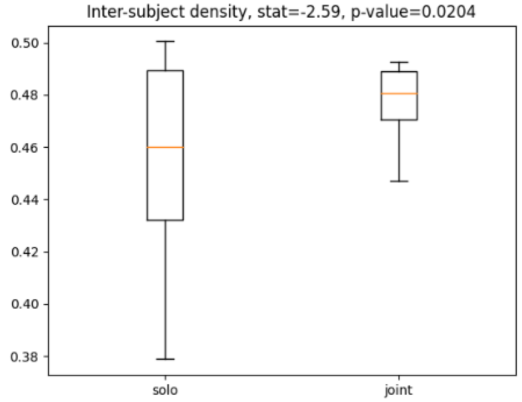


Results of Method 2 – Self-attention on Layer 1

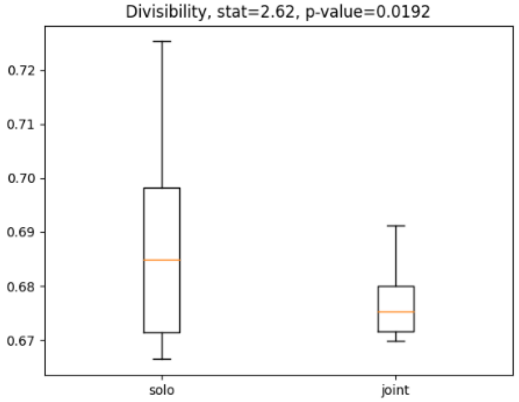
Inter-subject sum



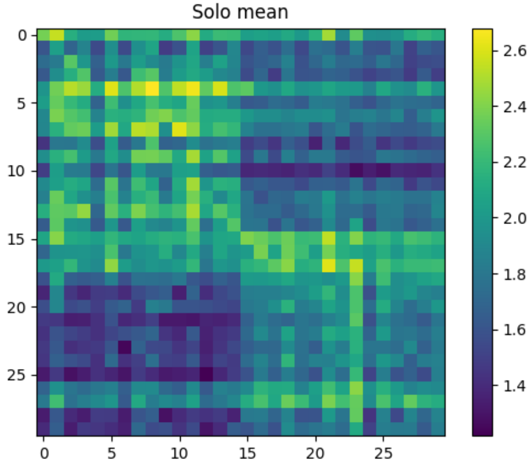
Inter-subject density



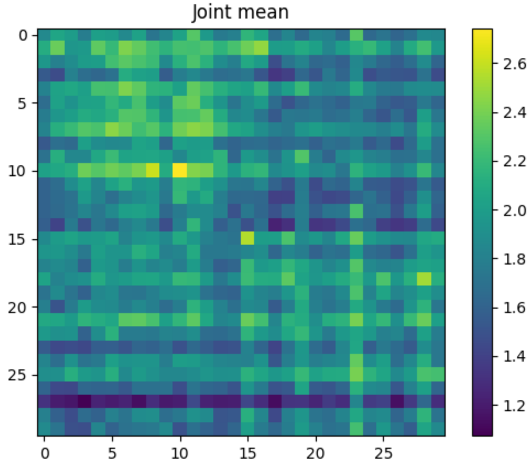
Divisibility



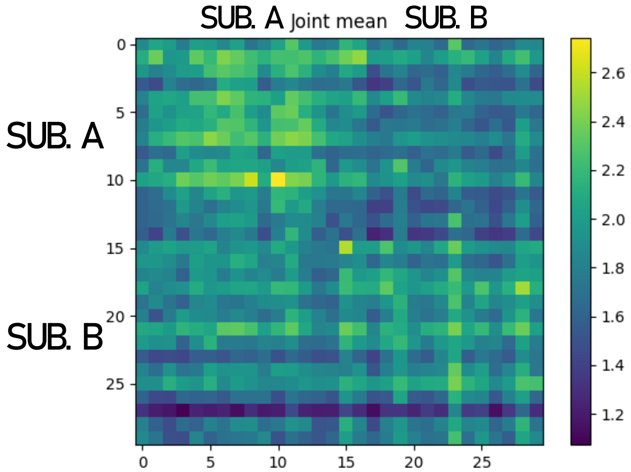
Solo condition



Joint condition

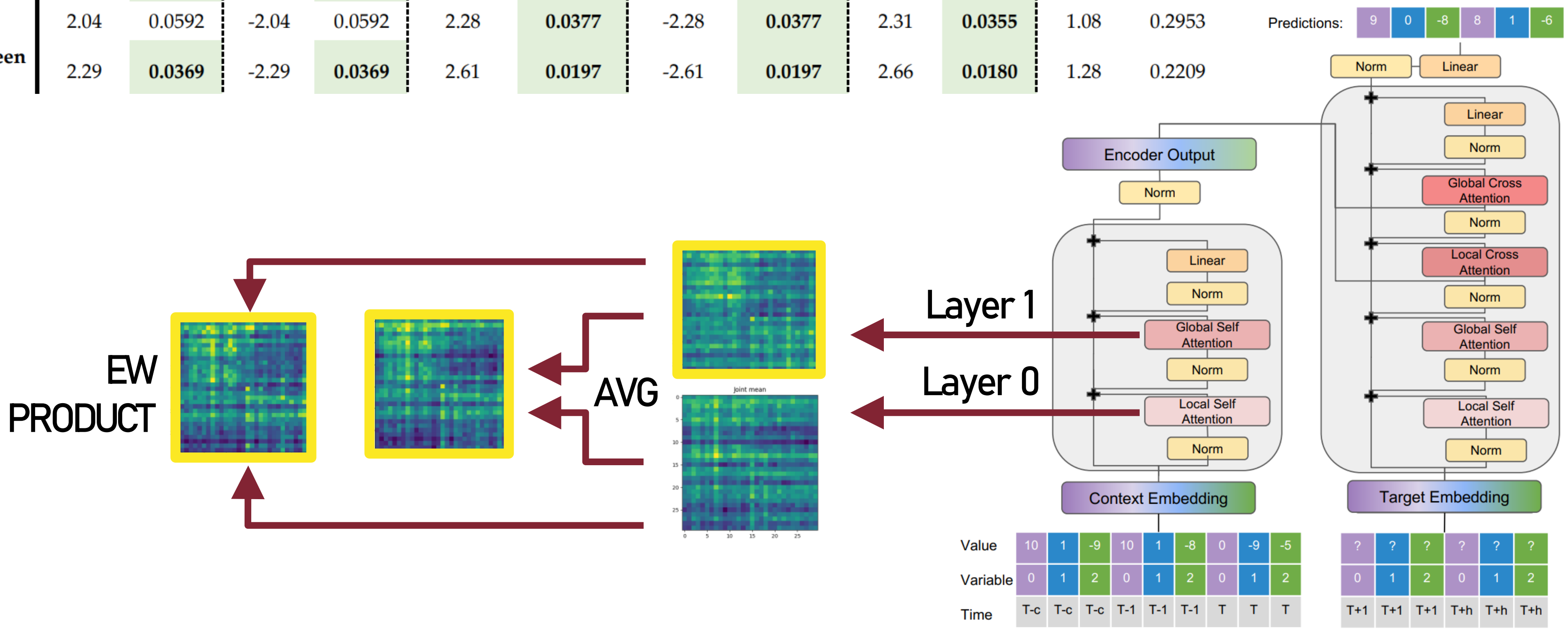


JOINT



Results of Method 2 – Significant types of aggregations

Matrix type	Sum intra-sub. Stat.	Sum intra-sub. P-value	Sum inter-sub. Stat.	Sum inter-sub. P-value	Density intra-sub. Stat.	Density intra-sub. P-value	Density inter-sub. Stat.	Density inter-sub. P-value	Divisibility Stat.	Divisibility P-value	Modularity Stat.	Modularity P-value
6. Self -attention on Layer 1	2.40	**0.0296**	-2.40	**0.0296**	2.59	**0.0204**	-2.59	**0.0204**	2.62	**0.0192**	1.64	0.1224
7. Average of self-attentions	2.04	0.0592	-2.04	0.0592	2.28	**0.0377**	-2.28	**0.0377**	2.31	**0.0355**	1.08	0.2953
8. Element-wise product between self-attentions	2.29	**0.0369**	-2.29	**0.0369**	2.61	**0.0197**	-2.61	**0.0197**	2.66	**0.0180**	1.28	0.2209



Conclusion

Method 1

Modify Conditioned Granger Causality with Spacetimeformer

- The **novel approach** showed performances **comparable to** those of the **Linear Conditioned Granger** causality **method**.
- **Long computational times**, (around 600 min on a Tesla T4 on a hyperscanning dyad).

Method 2

Use attention matrices as a causality measure

- The method provided physiologically **meaningful** and **sound results**, supported by **statistical evidence**.
- **Computational times** are sensibly **lower** (more then 30 times lower, 15-20 min on a Tesla T4 on a hyperscanning dyad).

Novelty and Future works

Novel aspects

I proposed two novel methods for the estimation of brain causality:

1. Base Conditioned Granger causality on Spacetimeformer residuals
2. Attention matrices as a measure of causality

First application of a Deep Learning method to EEG hyperscanning data.

Future research avenues

- Experimenting with **different aggregation** functions.
- Varying the **number** of **decoder** and **encoder** layers.
- **Correlating** computed indices with **behavioural data**.
- Using the **same** attention **method for** estimating **temporal relationships**.

Thank you

Luca Maurici 1809678

Master Thesis – M.Sc. in Artificial Intelligence and Robotics

25-10-2022

Relatore: Prof.ssa Laura Astolfi

Correlatore: Prof. Nicola Toschi

Controrelatore: Prof. Luca Iocchi