KU LEUVEN

STADIUS
Center for Dynamical Systems, Signal Processing and Data Analytics

Citation/Reference	Geirnaert S., Yao Y., Francart T., Bertrand A. (2025), **Stimulus-Informed Generalized Canonical Correlation Analysis for Group Analysis of Neural Responses to Natural Stimuli** *IEEE Journal of Biomedical and Health Informatics*, vol. 29, no. 2, pp. 970-983, Feb. 2025.
Archived version	Author manuscript: the content is identical to the content of the published paper, but without the final typesetting by the publisher
Published version	https://doi.org/10.1109/JBHI.2024.3462991
Journal homepage	https://www.embs.org/jbhi/
Author contact	simon.geirnaert@esat.kuleuven.be Klik hier als u tekst wilt invoeren.
Abstract	

(article begins on next page)

Stimulus-Informed Generalized Canonical Correlation Analysis for Group Analysis of Neural Responses to Natural Stimuli

Simon Geirnaert, Yuanyuan Yao, Tom Francart, and Alexander Bertrand, *Senior Member, IEEE*

***Abstract*— Various new brain-computer interface technologies or neuroscience applications require decoding stimulus-following neural responses to natural stimuli such as speech and video from, e.g., electroencephalography (EEG) signals. In this context, generalized canonical correlation analysis (GCCA) is often used as a group analysis technique, which allows the extraction of correlated signal components from the neural activity of multiple subjects attending to the same stimulus. GCCA can be used to improve the signal-to-noise ratio of the stimulus-following neural responses relative to all other irrelevant (non-)neural activity, or to quantify the correlated neural activity across multiple subjects in a group-wise coherence metric. However, the traditional GCCA technique is stimulus-unaware: no information about the stimulus is used to estimate the correlated components from the neural data of several subjects. Therefore, the GCCA technique might fail to extract relevant correlated signal components in practical situations where the amount of information is limited, for example, because of a limited amount of training data or group size. This motivates a new stimulus-informed GCCA (SI-GCCA) framework that allows taking the stimulus into account to extract the correlated components. We show that SI-GCCA outperforms GCCA in various practical settings, for both auditory and visual stimuli. Moreover, we showcase how SI-GCCA can be used to steer the estimation of the components towards the stimulus. As such, SI-GCCA substantially improves upon GCCA for various purposes, ranging from preprocessing to quantifying attention.**

***Index Terms*— generalized canonical correlation analysis, correlated component analysis, electroencephalography, stimulus-following neural response**

I. Introduction

Traditionally, brain-computer interface (BCI) and other neuroscience applications are oriented towards active paradigms, requiring the active participation of a user following instructions. Furthermore, they use multi-trial designs, requiring the same stimulus to be repeated multiple times to be able to average the brain responses to enhance the signal-to-noise (SNR) ratio. Moreover, synthetic stimuli such as flickering checkerboard patterns or beep sounds are used to elicit more controllable and deterministic neural responses, such as the P300-response or steady-state visual-evoked potentials (SSVEPs) [1]. While such controlled BCI paradigms are valuable from a scientific point of view, their practical impact is often limited to a few niche applications, for example, to re-establish communication for patients suffering from locked-in syndrome [1], [2]. To open up BCI technology to much more widespread usage in the daily life context, the limiting artificial conditions of such controlled BCI paradigms need to be alleviated.

In the past few years, we have seen a surge of BCI applications that are *passive*, i.e., tapping into the natural behavior of a user, *single-trial*, i.e., not requiring repetitions of the stimulus, and operate on *natural* sensory stimuli, such as speech/music and natural video [3]–[13]. These new BCI paradigms can then be employed in much more mainstream application domains, such as hearing aids and consumer earphones [3], [4], [14]–[16], educational sciences [5], [11], [17], neuromarketing [6], or virtual reality environments [18]. Many of these applications involve decoding stimulus-following neural responses, for example, to quantify levels of absolute [8], [10], [19] or selective attention [3], [14] to a particular auditory or visual stimulus. The temporal dynamics of the stimulus then result in a so-called stimulus-following neural response (i.e., the neural tracking phenomenon [20]), which can be decoded from different neurorecording modalities such as electroencephalography (EEG) [20], magnetoencephalography (MEG) [21], or electrocorticography (ECoG) [22]. The EEG modality is particularly interesting because it is non-invasive, low cost and highly mobile [1].

However, decoding neural responses to natural stimuli is much more challenging from a signal processing perspective, given that they are much more unpredictable (resulting from the nature of the stimulus) and suffer from a very low SNR, as the targeted stimulus-following neural responses are buried under all kinds of non-neural and neural noise. Given the single-trial paradigm, averaging the responses across multiple trials is not an option anymore to deal with this extremely low SNR. Therefore, much more advanced, data-driven signal processing algorithms are required and being developed to

This work is supported by a junior postdoctoral fellowship fundamental research from the Research Foundation Flanders (FWO) (for S. Geirnaert - 1242524N), FWO project nr. G081722N, the European Research Council (ERC) under the European Union's Horizon 2020 Research and Innovation Programme (grant agreement No 802895), and the Flemish Government (AI Research Program).

S. Geirnaert, Y. Yao, and A. Bertrand are with KU Leuven, Department of Electrical Engineering (ESAT), STADIUS Center for Dynamical Systems, Signal Processing and Data Analytics and with Leuven.AI - KU Leuven institute for AI, Kasteelpark Arenberg 10, B–3001 Leuven, Belgium (e-mail: simon.geirnaert@esat.kuleuven.be, yuanyuan.yao@esat.kuleuven.be, alexander.bertrand@esat.kuleuven.be).

T. Francart and S. Geirnaert are with KU Leuven, Department of Neurosciences, Research Group ExpORL, Herestraat 49 box 721, B–3000 Leuven, Belgium and with Leuven.AI - KU Leuven institute for AI (e-mail: tom.francart@kuleuven.be).

enhance the targeted stimulus-following activity and suppress all other noise [3], [23], [24]. In this paper, we focus on the *group analysis* of such stimulus-following responses, i.e., we assume a set of synchronized neural responses to the same natural stimulus is available, e.g., from a group of subjects all attending to the same stimulus. This synchronization hypothesis roots in the previously mentioned neural tracking phenomenon of natural stimuli, where the coherent neural responses across subjects arise from the synchronized stimulus-following neural responses. Importantly, this synchronization across subjects refers only to these stimulus-locked responses that are typically earlier responses, appearing in the 0–$400\,\mathrm{ms}$ post-stimulus range [14], [20]–[22]. More complex, longer-latency responses that are, for example, related to emotional arousal or cognition are unlikely to be synchronous across subjects. This hypothesis of synchronicity for group decoding has been confirmed in various papers in literature using data-driven decoding algorithms (e.g., [5]–[8], [13], [25], [26]). The group analysis of the stimulus-following neural responses can be a goal in itself, e.g., to decode a notion of group attention to the stimulus, as employed in educational neuroscience [5] or neuromarketing [6]. Alternatively, the group information can be leveraged to assist the decoding of stimulus-following neural responses on an individual level, using it as, e.g., a preprocessing technique to a priori improve the SNR [25], [26].

In this paper, we specifically focus on a signal processing technique called *generalized canonical correlation analysis* (GCCA), the multi-view extension of the frequently used canonical correlation analysis (CCA) [27], [28]. CCA extracts the correlated components between two views of the same activity [29]. In more traditional BCI's that, for example, decode SSVEPs, CCA is one of the most commonly used algorithms for classification by maximizing the correlation between the EEG responses and template reference signals that model the different flickering stimuli at their specific frequency (including harmonics) [30]. In the context of natural stimulus-following neural responses, CCA is, for example, used to decode the speech envelope of an attended speech source from EEG [9], [31]. In CCA, a decoder on the multi-channel EEG is then simultaneously trained with an encoder on the speech envelope to find these correlated components. The resulting correlation can then be used to decode selective attention, e.g., between two competing speech sources [3]. Its multi-view extension is GCCA, where the objective is to decode correlated components between more than two views of the same activity [27], [28]. In SSVEP-based BCIs, GCCA has been used to extract more natural reference signals by extracting the correlated components from several EEG trials containing responses to the same stimulus frequency. These learned reference signals can then be used in the previous two-view CCA method to classify new SSVEPs [32]. GCCA is moreover often used when multiple EEG signals from several users are available that simultaneously attend to the same natural stimulus. GCCA can then be used to quantify attention, enhance the SNR, reduce the dimensionality, or summarize the set of EEG signals [5], [6], [13], [25]. An excellent tutorial paper on GCCA (there dubbed MCCA) for decoding brain responses is written by de Cheveigné et al. [25].

A property of GCCA in this context is that it is *stimulus-unaware*: to extract the correlated components from the synchronized EEG signals, it does not assume or use any stimulus information. This is attractive in situations where the stimulus is unknown or unavailable, or when it is unknown what features of the stimulus elicit decodable neural responses. However, this stimulus-unawareness can be a disadvantage at the same time, e.g., when the stimulus is available or known, which occurs when a particular stimulus is deliberately used (e.g., in neuromarketing) or can be recorded (e.g., in hearing aids, in the classroom). Exploiting the stimulus in those situations as side information to help (e.g., in a regularization context) or steer the estimation of the correlated components across the synchronized EEG activity can then be highly beneficial, especially if we consider the very low SNR of the stimulus-following neural responses. This very low SNR is even harder to cope with when the amount of estimation data is limited, as is the case in a time-adaptive, online processing context [4], or when the group size is limited by the application. Therefore, the objective of this paper is to develop and analyze a *stimulus-informed* GCCA (SI-GCCA) algorithm that allows taking the stimulus into account when performing a group analysis of stimulus-following neural responses. While the use of CCA to extract correlated components between (individual) neural responses and the (natural) stimulus has been successfully established (see before and, e.g., [9], [30], [31]), such a group analysis of multiple stimulus-following neural responses where the neural decoders are specifically steered to yield responses that are (more) coherent with the natural auditory/visual stimulus has not yet been developed.

The paper is structured as follows. In Section II, we explain the well-known MAXVAR algorithm for GCCA and its correlated component analysis (corrCA) variant. In Section III, we derive our proposed SI-GCCA algorithm. In Section IV, we then describe all necessary details about the datasets and experiments to analyze the developed algorithms. The results are shown and discussed in Section V, and conclusions are drawn in Section VI.

Disclaimer: A conference precursor of this manuscript has been published in [33]. The current manuscript contains a more extensive explanation of the developed algorithm, includes an additional variant (e.g., the corrCA and SI-corrCA algorithm), does not only focus on speech as in [33] but also includes video, includes additional experiments, and provides additional use cases of the algorithm and a more in-depth analysis.

II. MAXVAR-GCCA/CORRCA

In canonical correlation analysis (CCA), the objective is to find the components that exhibit the highest correlation across two different views or subjects. There exist multiple options to generalize CCA to more than two views or subjects as envisaged in generalized canonical correlation analysis (GCCA), such as MAXVAR, SUMCORR, SSQCORR, … [27], [28] However, the two most popular variants are SUMCORR and MAXVAR, which both have the traditional CCA for two views as a special

case [34]. SUMCORR naturally extends CCA to more than two views by maximizing the sum of pairwise correlations between the different filtered views. However, this optimization problem turns out to be NP-hard with no closed-form solution [35]. Therefore, a relaxation of the SUMCORR-problem is often used, the MAXVAR-problem, in which the average pairwise distance between the filtered views is minimized. Hence, MAXVAR starts from a different interpretation of the CCA problem and conveniently boils down to a closed-form solution in the form of a generalized eigenvalue decomposition (GEVD), similar to CCA [27], [36], [37]. An interesting property of such a GEVD is that it is invariant under a scaling of one the views in the (G)CCA problem (see Lemma I in Hassani et al. [38]). Because of this attractive property of the MAXVAR-GCCA formulation (i.e., boiling down to a GEVD) and because it allows for an easier introduction of the stimulus information (see Section III), we choose this formulation in this work. The MAXVAR-GCCA formulation for stimulus-following neural responses is introduced in Section II-A. In the context of a group analysis of stimulus-following neural responses [25], sometimes an additional constraint is added to enforce identical neural decoders across all subjects [5], [6], [9]. GCCA with this additional constraint is often referred to as 'correlated component analysis' (corrCA) and will be briefly reviewed in Section II-B. The additional constraint in corrCA acts as a regularizer to limit the degrees of freedom in the model, avoiding overfitting in cases where the amount of data is limited.

A. MAXVAR-GCCA

We consider the EEG data from K subjects attending to the same natural stimulus, for example, a speech signal or video. This setup is visualized in Figure 1. We denote $\mathbf{X}_k \in \mathbb{R}^{T\times M}$ as the EEG data of the kth subject, where T denotes the number of available EEG samples. Each EEG sample is M-dimensional, corresponding to, for example, different EEG channels and/or time-lagged copies of each channel, where the latter allows to also exploit spectral or temporal information in the data-driven decoder design. Such temporal filtering is, for example, needed to compensate for temporal differences in neural processing between subjects in the group. Given C EEG channels and L time lags, ranging from $-\frac{L-1}{2}$ to $\frac{L-1}{2}$ (assuming L is odd), the EEG data matrix $\mathbf{X}_k$ when using spatiotemporal filtering will then consist of $M = CL$ columns. The resulting EEG regression matrix $\mathbf{X}_k$ is in that case a block Hankel matrix, e.g., for $L = 5$:

$$\mathbf{X}_k = \begin{bmatrix}\mathbf{X}_{k,1} & \dots & \mathbf{X}_{k,C}\end{bmatrix},$$

$$\mathbf{X}_{k,c} = \begin{bmatrix} 0 & 0 & x_{k,c}(0) & x_{k,c}(1) & x_{k,c}(2) \\ 0 & x_{k,c}(0) & x_{k,c}(1) & x_{k,c}(2) & x_{k,c}(3) \\ x_{k,c}(0) & x_{k,c}(1) & x_{k,c}(2) & x_{k,c}(3) & x_{k,c}(4) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{k,c}(T-3) & x_{k,c}(T-2) & x_{k,c}(T-1) & 0 & 0 \end{bmatrix},$$

with $x_{k,c}(t)$ the cth-channel EEG signal of the kth subject. While not strictly necessary, for simplicity, we assume an equal dimensionality per subject, i.e., M is the same across the EEG recordings of all subjects. We also assume that the EEG data in $\mathbf{X}_k$ is centered or high-pass filtered such that it is zero-mean.

The objective is to find the $M \times Q$-dimensional neural decoders $\mathbf{W}_k \in \mathbb{R}^{M\times Q}$ such that the individual projected signals $\mathbf{X}_k\mathbf{W}_k \in \mathbb{R}^{T\times Q}$ across all subjects $k = 1, \dots, K$ are on average as close as possible to each other. This can be realized by introducing a Q-dimensional shared signal subspace spanned by the columns of $\mathbf{S} = \begin{bmatrix}\mathbf{s}_1 & \cdots & \mathbf{s}_Q\end{bmatrix} \in \mathbb{R}^{T\times Q}$, and also optimizing for this shared signal subspace to be on average closest to $\mathbf{X}_k\mathbf{W}_k$, for all k [28], [37], [39]:

MAXVAR-GCCA problem

$$\begin{aligned} \min_{\mathbf{W}_1,\dots,\mathbf{W}_K,\mathbf{S}} \quad & \sum_{k=1}^{K} ||\mathbf{S} - \mathbf{X}_k\mathbf{W}_k||_F^2 + \mu \sum_{k=1}^{K} ||\mathbf{W}_k||_F^2 \\ \text{s.t.} \quad & \mathbf{S}^{\mathsf{T}}\mathbf{S} = \mathbf{I}_Q, \end{aligned} \tag{1}$$

with $\mathbf{I}_Q$ the Q-dimensional identity matrix and $||\cdot||_F$ the Frobenius norm. The second term in (1) represents ℓ_2-norm regularization or diagonal loading to avoid overfitting, where the hyperparameter μ controls the amount of regularization added. If $\mu = 0$, no regularization is used, corresponding to the original MAXVAR-GCCA problem. The constraint $\mathbf{S}^{\mathsf{T}}\mathbf{S} = \mathbf{I}_Q$ ensures that subsequent neural decoders have orthogonal/uncorrelated outputs *on average*[1] and avoids the trivial solution where all $\mathbf{W}_k$'s and $\mathbf{S}$ are set to zero.

Defining $\mathbf{R}_{kl} = \mathbf{X}_k^{\mathsf{T}}\mathbf{X}_l \in \mathbb{R}^{M\times M}$ as the sample crosscorrelation matrix of $\mathbf{X}_k$ and $\mathbf{X}_l$ (autocorrelation matrix when $k = l$),

$$\mathbf{R}_{D_{xx}} = \text{Blkdiag}(\mathbf{R}_{11}, \dots, \mathbf{R}_{KK}) \in \mathbb{R}^{KM\times KM}$$

represents the block diagonal matrix containing the per-subject autocorrelation matrices, and

$$\mathbf{R}_{xx} = \mathbf{X}^{\mathsf{T}}\mathbf{X} \in \mathbb{R}^{KM\times KM},$$

with $\mathbf{X} = \begin{bmatrix}\mathbf{X}_1 & \cdots & \mathbf{X}_K\end{bmatrix} \in \mathbb{R}^{T\times KM}$, is the correlation matrix of all EEG data, containing all correlation matrices $\mathbf{R}_{kl}$ in its blocks. The Karush-Kuhn-Tucker (KKT) conditions then lead to the following generalized eigenvalue problem[2] [37], [39]:

MAXVAR-GCCA solution

$$(\mathbf{R}_{D_{xx}} + \mu\mathbf{I}_{KM})\,\mathbf{W} = \mathbf{R}_{xx}\mathbf{W}\mathbf{\Omega}, \tag{2}$$

where $\mathbf{W} = \begin{bmatrix}\mathbf{W}_1^{\mathsf{T}} & \cdots & \mathbf{W}_K^{\mathsf{T}}\end{bmatrix}^{\mathsf{T}} \in \mathbb{R}^{KM\times Q}$ concatenates all per-subject neural decoders. We find the optimal neural decoders $\mathbf{W}$ as the Q generalized eigenvectors (GEVcs) corresponding to the Q smallest generalized eigenvalues (GEVls), which can be found in the diagonal matrix $\mathbf{\Omega} \in \mathbb{R}^{Q\times Q}$ (see [37] and the derivation of SI-GCCA in Section III-A).

[1] This is a major difference with the SUMCORR formulation, where the subsequent neural decoders for each *individual* subjects should yield uncorrelated outputs.

[2] The derivation of (2) is a special case of the derivation of the SI-GCCA solution in Section III-A, i.e., for $\gamma = 0$.

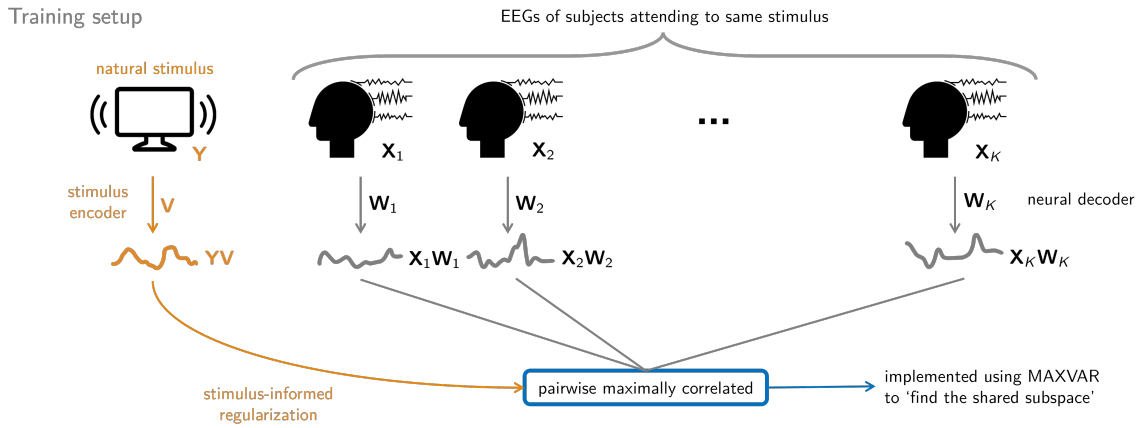

Fig. 1: In this work, we consider the EEG data $\mathbf{X}_k$ of K subjects attending to the same natural stimulus, represented by $\mathbf{Y}$. In GCCA, the intuitive objective is to find per-subject neural decoders $\mathbf{W}_k$ that maximize the pairwise correlation between the neurally decoded signals $\mathbf{X}_k\mathbf{W}_k$. Here we implement this objective via the MAXVAR framework. In our proposed SI-GCCA framework, the stimulus is included in the optimization problem via a stimulus encoder $\mathbf{V}$ to steer and regularize the estimation problem.

Furthermore, it can be found that

$$\mathbf{S} = \sum_{k=1}^{K} \mathbf{X}_k\mathbf{W}_k\mathbf{\Omega}. \tag{3}$$

The correct scaling of the GEVcs and thus neural decoders is determined via the equality constraint $\mathbf{S}^\mathsf{T}\mathbf{S} = \mathbf{I}_Q$ and (3).

Note that alternatively to solving the generalized eigenvalue decomposition (GEVD) in (2), the neural decoders can also be found via the eigenvalue decomposition of the correlation matrix of all the pre-whitened data matrices of each subject. As such, GCCA can be formulated as concatenating two principal component analysis (PCA) blocks in a two-step analysis, which might be more intuitive and easy to understand, as explained and motivated in [25]. As furthermore explained in [25], the inverse eigenvalues of (2) represent the degree of correlation that exists for each component. A component that is only present in the data of a single subject corresponds to a unit eigenvalue, whereas a component shared by exactly P data matrices corresponds to an associated eigenvalue of $\frac{1}{P}$. As such, the eigenvalues are smaller for signal components that are shared by many subjects.

B. MAXVAR-corrCA

In the GCCA problem in Section II-A, for each subject k, a different neural decoder $\mathbf{W}_k$ is trained. As an additional constraint, one could restrict these per-subject neural decoders to be the same across subjects, i.e., $\mathbf{W}_1 = \cdots = \mathbf{W}_K = \mathbf{W} \in \mathbb{R}^{M\times Q}$, which is dubbed corrCA in [5], [6], [9]. From a parameter estimation point of view, this can be viewed as an additional regularization, as the number of parameters that need to be estimated drastically decreases. From a neural point of view, this assumes more uniform neural signals across subjects.

In MAXVAR-corrCA with diagonal loading, the optimization problem in (1), therefore, boils down to:

MAXVAR-corrCA problem

$$\begin{aligned} \min_{\mathbf{W},\mathbf{S}} \quad & \sum_{k=1}^{K} ||\mathbf{S} - \mathbf{X}_k\mathbf{W}||_F^2 + \mu\,||\mathbf{W}||_F^2 \\ \text{s.t.} \quad & \mathbf{S}^\mathsf{T}\mathbf{S} = \mathbf{I}_Q. \end{aligned} \tag{4}$$

The KKT conditions then lead to the following generalized eigenvalue problem[3]:

MAXVAR-corrCA solution

$$\left(\sum_{k=1}^{K} \mathbf{R}_{kk} + \mu\mathbf{I}_M\right)\mathbf{W} = \left(\sum_{k=1}^{K}\sum_{l=1}^{K} \mathbf{R}_{kl}\right)\mathbf{W}\mathbf{\Omega}. \tag{5}$$

As opposed to the GEVD in (2), which is of dimension KM, this GEVD is only of dimension M, effectively showing the regularization aspect of corrCA. Moreover, the matrices in (5) correspond to the sum of the $M \times M$ block matrices of $\mathbf{R}_{D_{xx}}$ and $\mathbf{R}_{xx}$, respectively, showing the effect of restricting the per-subject neural decoders to be the same.

III. SI-GCCA/SI-corrCA

We now propose the stimulus-informed GCCA (SI-GCCA) technique for the group analysis of stimulus-following neural responses, in which we include the stimulus as side information, as visualized in Figure 1. The implicit assumption in SI-GCCA is, therefore, that the correlated components across the K different EEG recordings correspond to (early) stimulus-following/-related neural responses, as explained in the introduction. This assumption also delineates the context in which we propose SI-GCCA, i.e., all K EEG recordings are recorded using the same natural stimulus and are synchronized in time.

The stimulus is incorporated in the GCCA technique with two objectives in mind. Firstly, to steer the estimation of

[3] Again, the derivation can be viewed as a special case of the derivation in Section III-A, when $\gamma = 0$.

the correlated components towards the stimulus and specific stimulus representations. Secondly, it can be viewed as task-informed regularization (in the broad sense), as it allows to take additional information into account to cope with situations where less information (e.g., little data, few subjects) is available. Both objectives will be evaluated in Section V.

A. SI-GCCA

Apart from the K zero-mean EEG signals $\mathbf{X}_k$, now also assume we have access to the stimulus via the P-dimensional stimulus representation $\mathbf{Y} \in \mathbb{R}^{T\times P}$ (see Section IV-B.1 for examples of such a representation for a speech and video stimulus). In SI-GCCA, our goal is to ensure that the shared signal subspace $\mathbf{S}$, which connects the EEGs of the different subjects, is also close to the stimulus representation $\mathbf{Y}$, i.e., the signals in $\mathbf{S}$ and $\mathbf{Y}$ should be correlated. We do this by introducing an extra term into the MAXVAR-GCCA estimation problem in (1), where we use a forward model/encoder $\mathbf{V} \in \mathbb{R}^{P\times Q}$ on the stimulus to map it to the shared signal subspace $\mathbf{S}$:

SI-GCCA problem

$$\begin{aligned}\min_{\substack{\mathbf{W}_1,\ldots,\mathbf{W}_K,\\ \mathbf{V},\mathbf{S}}} \quad & \sum_{k=1}^{K} ||\mathbf{S}-\mathbf{X}_k\mathbf{W}_k||_F^2 + \gamma\,||\mathbf{S}-\mathbf{Y}\mathbf{V}||_F^2 \\ & + \mu\left(\sum_{k=1}^{K}||\mathbf{W}_k||_F^2 + ||\mathbf{V}||_F^2\right) \\ \text{s.t.} \quad & \mathbf{S}^{\mathrm{T}}\mathbf{S} = \mathbf{I}_Q.\end{aligned} \tag{6}$$

The hyperparameter γ determines how much weight is put onto the stimulus, i.e., how hard the stimulus 'pulls' on the shared signal subspace. The motivation behind using a simple forward model $\mathbf{V}$ on the stimulus representation is that it essentially retains the MAXVAR-GCCA structure as in (1), where the stimulus acts as an additional view of the underlying shared subspace. We will show that, as a result, SI-GCCA retains the attractive property of MAXVAR-GCCA that it can be solved via a GEVD.

The Lagrangian function is equal to:

$$\begin{aligned}\mathcal{L}(\mathbf{W}_1,\ldots,\mathbf{W}_K,\mathbf{V},\mathbf{S},\mathbf{\Lambda}) = {} & (K+\gamma)\mathrm{Tr}\left(\mathbf{S}^{\mathrm{T}}\mathbf{S}\right) - \\ & 2\sum_{k=1}^{K}\mathrm{Tr}\left(\mathbf{S}^{\mathrm{T}}\mathbf{X}_k\mathbf{W}_k\right) + \sum_{k=1}^{K}\mathrm{Tr}\left(\mathbf{W}_k^{\mathrm{T}}\mathbf{X}_k^{\mathrm{T}}\mathbf{X}_k\mathbf{W}_k\right) - \\ & 2\gamma\mathrm{Tr}\left(\mathbf{S}^{\mathrm{T}}\mathbf{Y}\mathbf{V}\right) + \gamma\mathrm{Tr}\left(\mathbf{V}^{\mathrm{T}}\mathbf{Y}^{\mathrm{T}}\mathbf{Y}\mathbf{V}\right) + \mu\sum_{k=1}^{K}\mathrm{Tr}\left(\mathbf{W}_k^{\mathrm{T}}\mathbf{W}_k\right) + \\ & \mu\mathrm{Tr}\left(\mathbf{V}^{\mathrm{T}}\mathbf{V}\right) - \mathrm{Tr}\left(\left(\mathbf{S}^{\mathrm{T}}\mathbf{S}-\mathbf{I}_Q\right)\mathbf{\Lambda}\right),\end{aligned}$$

with $\mathbf{\Lambda} \in \mathbb{R}^{Q\times Q}$ a symmetric matrix containing the Lagrange multipliers. The KKT conditions then lead to the following four equations:

$$\begin{cases} \nabla_{\mathbf{W}_k}(\mathcal{L}) = 0 \Leftrightarrow \mathbf{X}_k^{\mathrm{T}}\mathbf{S} = \left(\mathbf{R}_{kk}+\mu\mathbf{I}_M\right)\mathbf{W}_k,\ \forall k & (7)\\ \nabla_{\mathbf{V}}(\mathcal{L}) = 0 \Leftrightarrow \gamma\mathbf{Y}^{\mathrm{T}}\mathbf{S} = \left(\gamma\mathbf{R}_{yy}+\mu\mathbf{I}_P\right)\mathbf{V}, & (8)\\ \nabla_{\mathbf{S}}(\mathcal{L}) = 0 \Leftrightarrow \mathbf{S} = \left(\sum_{k=1}^{K}\mathbf{X}_k\mathbf{W}_k+\gamma\mathbf{Y}\mathbf{V}\right)\mathbf{\Omega}, & (9)\\ \nabla_{\mathbf{\Lambda}}(\mathcal{L}) = 0 \Leftrightarrow \mathbf{S}^{\mathrm{T}}\mathbf{S} = \mathbf{I}_Q, & (10)\end{cases}$$

with $\mathbf{R}_{yy} = \mathbf{Y}^{\mathrm{T}}\mathbf{Y} \in \mathbb{R}^{P\times P}$ the stimulus autocorrelation matrix and $\mathbf{\Omega} = \left((K+\gamma)\mathbf{I}_Q - \mathbf{\Lambda}\right)^{-1} \in \mathbb{R}^{Q\times Q}$ a symmetric matrix. Define the crosscorrelation matrix between EEG data matrix $\mathbf{X}_k$ and stimulus data matrix $\mathbf{Y}$ as $\mathbf{R}_{ky} = \mathbf{X}_k^{\mathrm{T}}\mathbf{Y} \in \mathbb{R}^{M\times P}$, and the augmented data matrix $\bar{\mathbf{X}} = \begin{bmatrix}\mathbf{X}_1 & \cdots & \mathbf{X}_K & \mathbf{Y}\end{bmatrix} \in \mathbb{R}^{T\times(KM+P)}$ and variables $\bar{\mathbf{W}} = \begin{bmatrix}\mathbf{W}_1^{\mathrm{T}} & \cdots & \mathbf{W}_K^{\mathrm{T}} & \mathbf{V}^{\mathrm{T}}\end{bmatrix}^{\mathrm{T}} \in \mathbb{R}^{(KM+P)\times Q}$, both including the stimulus data matrix and, respectively, the forward encoder. By plugging (9) into (7) and (8), and combining all equations, we find:

SI-GCCA solution

$$\left(\mathbf{P}\mathbf{R}_{D_{\bar{x}\bar{x}}}+\mu\mathbf{I}_{KM+P}\right)\bar{\mathbf{W}} = \mathbf{P}\mathbf{R}_{\bar{x}\bar{x}}\mathbf{P}\bar{\mathbf{W}}\mathbf{\Omega}, \tag{11}$$

where $\mathbf{R}_{\bar{x}\bar{x}} = \bar{\mathbf{X}}^{\mathrm{T}}\bar{\mathbf{X}} \in \mathbb{R}^{(KM+P)\times(KM+P)}$, $\mathbf{R}_{D_{\bar{x}\bar{x}}} = \mathrm{Blkdiag}(\mathbf{R}_{11},\ldots,\mathbf{R}_{KK},\mathbf{R}_{yy})$, and

$$\mathbf{P} = \begin{bmatrix}\mathbf{I}_{KM} & \mathbf{0}\\ \mathbf{0} & \gamma\mathbf{I}_P\end{bmatrix} \text{(weighting matrix)}.$$

Using this in (11), we find

$$\mathbf{P}\mathbf{R}_{D_{\bar{x}\bar{x}}} = \mathrm{Blkdiag}\left(\mathbf{R}_{11},\ldots,\mathbf{R}_{KK},\gamma\mathbf{R}_{yy}\right),$$

and

$$\mathbf{P}\mathbf{R}_{\bar{x}\bar{x}}\mathbf{P} = \begin{bmatrix}\mathbf{R}_{11} & \ldots & \mathbf{R}_{1K} & \gamma\mathbf{R}_{1y}\\ \vdots & & \vdots & \vdots\\ \mathbf{R}_{K1} & \ldots & \mathbf{R}_{KK} & \gamma\mathbf{R}_{Ky}\\ \gamma\mathbf{R}_{y1} & \ldots & \gamma\mathbf{R}_{yK} & \gamma^2\mathbf{R}_{yy}\end{bmatrix}.$$

While (11) resembles a generalized eigenvalue problem, $\mathbf{\Omega}$ is not a diagonal matrix. However, it can be easily found that the underlying solution boils down to a GEVD, given that $\mathbf{\Omega}$ is orthogonally diagonalizable as it is a symmetric matrix:

$$\mathbf{\Omega} = \mathbf{U}\mathbf{\Sigma}\mathbf{U}^{\mathrm{T}}, \tag{12}$$

with $\mathbf{U} \in \mathbb{R}^{Q\times Q}$ an orthogonal matrix and $\mathbf{\Sigma} \in \mathbb{R}^{Q\times Q}$ a diagonal matrix. Substituting (12) in (11) reveals the underlying GEVD that leads to the solution:

$$\left(\mathbf{P}\mathbf{R}_{D_{\bar{x}\bar{x}}}+\mu\mathbf{I}_{KM+P}\right)\bar{\mathbf{W}}\mathbf{U} = \mathbf{P}\mathbf{R}_{\bar{x}\bar{x}}\mathbf{P}\bar{\mathbf{W}}\mathbf{U}\mathbf{\Sigma}. \tag{13}$$

From (13), it can be seen that the set of optimal neural decoders and stimulus encoders defined by $\bar{\mathbf{W}}$ are in the subspace spanned by the GEVcs from the matrix pencil $(\mathbf{P}\mathbf{R}_{D_{\bar{x}\bar{x}}}+\mu\mathbf{I}_{KM+P}, \mathbf{P}\mathbf{R}_{\bar{x}\bar{x}}\mathbf{P})$. Given that $\mathbf{P}$ is a symmetric matrix, and $\left(\mathbf{P}\mathbf{R}_{D_{\bar{x}\bar{x}}}\right)^{\mathrm{T}} = \mathbf{R}_{D_{\bar{x}\bar{x}}}^{\mathrm{T}}\mathbf{P}^{\mathrm{T}} = \mathbf{R}_{D_{\bar{x}\bar{x}}}\mathbf{P} = \mathbf{P}\mathbf{R}_{D_{\bar{x}\bar{x}}}$, (13) is the GEVD of a matrix pencil of two real *symmetric* matrices with $\mathbf{P}\mathbf{R}_{\bar{x}\bar{x}}\mathbf{P}$ positive definite, therefore, resulting in a real solution for GEVcs and GEVls (found on the diagonal of $\mathbf{\Sigma}$). Furthermore, as the solution of (6) is only defined upon any orthogonal transformation, and since the objective function at the optimal solution can be found[4] to be equal to $(K+\gamma)Q - \mathrm{Tr}\left(\mathbf{\Sigma}^{-1}\right)$, we find the optimal neural decoders $\mathbf{W}_k$ and stimulus encoder $\mathbf{V}$ as the Q GEVcs corresponding to the Q *smallest* GEVls. The equality constraint in (10) determines the correct scaling of the GEVcs. We summarize the SI-GCCA algorithm in Algorithm 1. A MATLAB implementation is available [40].

[4]Proofs are omitted for conciseness.

Algorithm 1: SI-GCCA

Input: K stimulus-driven EEG signals $\mathbf{X}_k \in \mathbb{R}^{T\times M}$, stimulus features $\mathbf{Y} \in \mathbb{R}^{T\times P}$, stimulus hyperparameter γ, ℓ_2-norm regularization hyperparameter μ, subspace dimension Q
Output: Per-subject neural decoders $\mathbf{W}_k \in \mathbb{R}^{M\times Q}$

1: Compute correlation matrices $\mathbf{R}_{\bar{x}\bar{x}}$ and $\mathbf{R}_{D_{\bar{x}\bar{x}}} = \text{Blkdiag}(\mathbf{R}_{11}, \ldots, \mathbf{R}_{KK}, \mathbf{R}_{yy})$, with $\bar{\mathbf{X}} = \begin{bmatrix}\mathbf{X}_1 & \cdots & \mathbf{X}_K & \mathbf{Y}\end{bmatrix}$, $\mathbf{R}_{kk} = \mathbf{X}_k^\mathsf{T}\mathbf{X}_k$, and $\mathbf{R}_{yy} = \mathbf{Y}^\mathsf{T}\mathbf{Y}$
2: Compute $\bar{\mathbf{W}}$ as the Q GEVcs corresponding to the Q smallest GEVls of the matrix pencil $(\mathbf{P}\mathbf{R}_{D_{\bar{x}\bar{x}}} + \mu\mathbf{I}_{KM+P}, \mathbf{P}\mathbf{R}_{\bar{x}\bar{x}}\mathbf{P})$, with $\mathbf{P} = \begin{bmatrix}\mathbf{I}_{KM} & \mathbf{0}\\ \mathbf{0} & \gamma\mathbf{I}_P\end{bmatrix}$
3: Scale the GEVcs such that $\mathbf{S}^\mathsf{T}\mathbf{S} = \mathbf{I}_Q$, with $\mathbf{S}$ defined in (9)
4: Extract $\mathbf{W}_k$ from $\tilde{\mathbf{W}} = \begin{bmatrix}\mathbf{W}_1^\mathsf{T} & \cdots & \mathbf{W}_K^\mathsf{T} & \mathbf{V}^\mathsf{T}\end{bmatrix}^\mathsf{T}$

B. SI-corrCA

In the stimulus-informed corrCA (SI-corrCA) version we constrain the per-subject neural decoders to be the same across subjects. In this scenario, we are most heavily introducing additional constraints into the group decoding problem: all neural decoders are the same across subjects (corrCA) *and* the correlated components must resemble the stimulus features (stimulus-informed). The optimization problem then becomes:

SI-corrCA problem

$$\begin{aligned}\min_{\mathbf{W},\mathbf{V},\mathbf{S}} \quad & \sum_{k=1}^{K} ||\mathbf{S} - \mathbf{X}_k\mathbf{W}||_F^2 + \gamma\,||\mathbf{S} - \mathbf{Y}\mathbf{V}||_F^2 \\ & + \mu\left(||\mathbf{W}||_F^2 + ||\mathbf{V}||_F^2\right) \\ \text{s.t.} \quad & \mathbf{S}^\mathsf{T}\mathbf{S} = \mathbf{I}_Q.\end{aligned} \tag{14}$$

The KKT conditions then, similarly to Section III-A, lead to the following generalized eigenvalue problem:

SI-corrCA solution

$$\left(\begin{bmatrix}\sum_{k=1}^{K}\mathbf{R}_{kk} & \mathbf{0}\\ \mathbf{0} & \gamma\mathbf{R}_{yy}\end{bmatrix} + \mu\mathbf{I}_{M+P}\right)\begin{bmatrix}\mathbf{W}\\ \mathbf{V}\end{bmatrix} = \begin{bmatrix}\sum_{k=1}^{K}\sum_{l=1}^{K}\mathbf{R}_{kl} & \gamma\sum_{k=1}^{K}\mathbf{R}_{ky}\\ \gamma\sum_{k=1}^{K}\mathbf{R}_{yk} & \gamma^2\mathbf{R}_{yy}\end{bmatrix}\begin{bmatrix}\mathbf{W}\\ \mathbf{V}\end{bmatrix}\mathbf{\Omega}, \tag{15}$$

merging properties from both the MAXVAR-corrCA solution in (5) and the SI-GCCA solution in (13).

IV. Experiments

We will evaluate and compare the different GCCA and SI-GCCA variants based on the EEG signals of a group of subjects listening to the same speech signals or watching the same videos. In this section, we describe the experiments in terms of the datasets (Section IV-A), the EEG preprocessing and stimulus feature extraction (Section IV-B), the performance evaluation schemes and metrics (Section IV-C), and the decoder setup (Section IV-D). MATLAB code to reproduce all experiments is available online [40].

A. Datasets

Two datasets are used to compare the various methods: one with natural speech and one with video footage as the stimulus.

1) Speech dataset: The speech dataset is taken from the first experiment of Broderick et al. [41] and contains the EEG data of 19 normal-hearing subjects listening to the same audiobooks. The EEG data is recorded with a 128-channel BioSemi ActiveTwo system and is re-referenced to the average mastoid channel. The data was recorded per subject separately in 20 trials of around 3 min long and is cut into 52 1 min-trials for further processing. In total, there is 52 min of synchronized EEG/speech data per subject. This dataset is publicly available [42].

2) Video dataset: The video dataset is taken from the Single-Shot dataset of Yao et al. [13] and contains the EEG data of 20 healthy subjects with normal or corrected-to-normal vision watching the same video footage. The video footage consists of a single moving person during a performance (e.g., dance, magic shows). The EEG data is recorded with a 64-channel BioSemi ActiveTwo system. The data was recorded per subject separately in 2 trials of around 36 and 35 min and is cut into 56 1 min-trials for further processing. This dataset is publicly available [43].

B. Stimulus feature extraction and EEG preprocessing

1) Stimulus features:

a) Speech: The speech signals are represented by the low-frequency envelope of the speech signal, computed using the Hilbert transform [44]. In various works (e.g., [14]–[16], [44]), it has been shown that the EEG signals track this speech envelope. Moreover, we bandpass-filter the speech envelope using a 4th-order Butterworth filter in the δ-band (1–4 Hz), where it is shown to give good tracking results [10]. The resulting signal is stored in a one-dimensional vector $\mathbf{y} \in \mathbb{R}^T$ containing the samples of the envelope at different time instances during the experiment. This representation $\mathbf{y}$ of the speech stimulus will be used in the SI-GCCA framework to create the matrix $\mathbf{Y}$ (see also Section IV-D).

Note that while in this work, we choose the speech envelope as an exemplary stimulus representation for natural speech, other (higher-level) features such as, e.g., phoneme and word onsets, phoneme and word surprisal, or cohort entropy could be (even additionally) used, as they are shown to also synchronize with neural signals [45]–[47].

b) Video: The video stimulus is represented by an object-based version of the average optical flow, i.e., the magnitude of the pixel-wise velocity vector between frames averaged across all pixels belonging to an object in the video (after object segmentation) [13]. This again results in a one-dimensional vector $\mathbf{y} \in \mathbb{R}^T$. In Yao et al. [13], it is shown that the object-based optical flow leads to significant tracking in the EEG

signals of subjects watching a video. This object-based optical flow is computed after resampling (including anti-aliasing) the video data to 30 Hz and resizing it to 854×480 pixels [13].

2) EEG preprocessing: The EEG data are preprocessed similarly to the original references of the datasets [13], [41]. This means that the EEG data for the video dataset is first preprocessed by interpolating bad channels, average re-referencing, notch filtering to remove the powerline noise, and regressing out eye activity using EOG. For the speech dataset, the EEG data were re-referenced to the average of the mastoid channels. Additionally, the EEG data is bandpass-filtered (between 1–4 Hz for the speech dataset and 0.5–15 Hz for the video dataset), downsampled (to 8 Hz for the speech dataset and 30 Hz for the video dataset), and normalized. In both cases, the EEG data are afterwards normalized per 1 min-trial by setting the mean per channel to zero and the Frobenius norm across all channels to one.

C. Performance evaluation

1) Testing procedure: To investigate the influence of different variables such as group size, amount of training data, and number of channels, we perform Monte-Carlo experiments in which we fix 2 variables to a default value, and perform a sweep on the third one. 50 Monte Carlo runs are used for each value in the sweep. Possible interactions between these 3 variables will be investigated ad hoc (see Section V-D). The default values of the different variables are 40 min of training data, 64 channels, and using all subjects. In the speech case, this means that the 64 channels corresponding to the 64-channel BioSemi system of the video dataset are chosen from the 128-channel EEG system of the speech dataset. As such, the baseline values between the speech and video datasets are the same. When 40 min of training are selected, the rest of the trials are split into 25% validation set (for hyperparameter estimation, see Section IV-D.2) and 75% test set. Per Monte Carlo run, the training trials are randomly sampled from all available trials. When varying the group size or number of channels, the chosen group or channels are similarly randomly changed between Monte Carlo runs. The same training-validation-test set split per Monte Carlo run is used for all methods, such that the results are directly comparable between methods.

The test window length, over which the correlation performance metrics below are computed, is 60 s.

2) Evaluation metrics: We consider two different evaluation metrics to compare the different methods: the inter-subject correlation (ISC) and stimulus correlation (SC). All evaluation metrics are schematically explained in Figure 2.

a) Inter-subject correlation (ISC): GCCA is often used to quantify the group attention to a specific natural stimulus by using the ISC as a proxy for attentional engagement [5]–[8], [13]. This ISC is defined as the average pairwise correlation coefficient between the GCCA-decoded EEG recordings of the different subjects, where the Pearson correlation coefficient $\rho(\mathbf{x}, \mathbf{y})$ between two zero-mean one-dimensional time signals $\mathbf{x} \in \mathbb{R}^T$ and $\mathbf{y} \in \mathbb{R}^T$ is defined as:

$$\rho(\mathbf{x}, \mathbf{y}) = \frac{\mathbf{x}^{\mathrm{T}}\mathbf{y}}{||\mathbf{x}||_2 \, ||\mathbf{y}||_2}.$$

Given the zero-mean one-dimensional projected EEG signals $\mathbf{z}_k^{(q)} = \mathbf{X}_k \mathbf{w}_k^{(q)} \in \mathbb{R}^T$ where $\mathbf{w}_k^{(q)}$ denotes the qth column of the (SI-)GCCA decoder $\mathbf{W}_k$ (or of $\mathbf{W}$ in the case of corrCA), the ISC for component q is thus defined as:

$$\mathrm{ISC}^{(q)} = \frac{2}{K(K-1)} \sum_{k=1}^{K-1} \sum_{l=k+1}^{K} \rho\left(\mathbf{z}_k^{(q)}, \mathbf{z}_l^{(q)}\right). \tag{16}$$

This ISC can be evaluated per window of a certain length on the test set to compare the various methods. Note that even for the stimulus-informed algorithms, at test time, we do not take the stimulus into account but compare them with the stimulus-unaware versions on exactly the same basis by only taking the projected EEG signals $\mathbf{z}_k^{(q)}$ into account.

b) Stimulus correlation (SC): Similarly to [25], [26], we can also use (SI-)GCCA/corrCA as a preprocessing tool to a priori enhance the SNR of the stimulus-following responses by leveraging the group information, to then perform a traditional backward stimulus correlation analysis on the projected EEG signals $\mathbf{Z}_k = \mathbf{X}_k \mathbf{W}_k \in \mathbb{R}^{T \times Q}$. We can assume the stimulus feature $\mathbf{y} \in \mathbb{R}^T$ to be one-dimensional here (see Section IV-B.1). The SC can then be found by first training a least-squares decoder $\mathbf{d}_k \in \mathbb{R}^{QL_d}$ from the time-lagged (using L_d additional time lags) projected EEG signals $\tilde{\mathbf{Z}}_k \in \mathbb{R}^{T \times QL_d}$ to the stimulus $\mathbf{y}$. As explained in [15], this decoder can be found by solving the normal equations:

$$\mathbf{d}_k = \left(\tilde{\mathbf{Z}}_k^{\mathrm{T}} \tilde{\mathbf{Z}}_k\right)^{-1} \tilde{\mathbf{Z}}_k^{\mathrm{T}} \mathbf{y}. \tag{17}$$

The SC for subject k can then be evaluated per window of a certain length on the test set by first applying the (SI-)GCCA decoders as a dimensionality reduction/preprocessing step and then applying the stimulus decoder:

$$\mathrm{SC}_k = \rho\left(\mathbf{y}, \tilde{\mathbf{Z}}_k \mathbf{d}_k\right). \tag{18}$$

While (18) contains the per-subject SC_k, we can also evaluate the overall SC across all subjects based on the subspace signal $\mathbf{S} = \frac{1}{K} \sum_{k=1}^{K} \mathbf{X}_k \mathbf{W}_k$ defined as the average projected EEG signal (similar to (3)), which then acts as the summary signal. Using a decoder

$$\mathbf{d}_{\text{avg}} = \left(\tilde{\mathbf{S}}^{\mathrm{T}} \tilde{\mathbf{S}}\right)^{-1} \tilde{\mathbf{S}}^{\mathrm{T}} \mathbf{y} \tag{19}$$

trained on the time-lagged average subspace signal $\tilde{\mathbf{S}} \in \mathbb{R}^{T \times QL_d}$ of the training set, the resulting stimulus correlation SC_{avg} then represents a proxy for the group attention:

$$\mathrm{SC}_{\text{avg}} = \rho\left(\mathbf{y}, \tilde{\mathbf{S}} \mathbf{d}_{\text{avg}}\right). \tag{20}$$

These stimulus decoders (17) and (19) are trained on the same data (i.e., the training set) on which the (SI-)GCCA/corrCA decoders are trained. That means the (SI-)GCCA/corrCA decoders are first applied to the training set, after which the decoders in (17) and (19) can be trained.

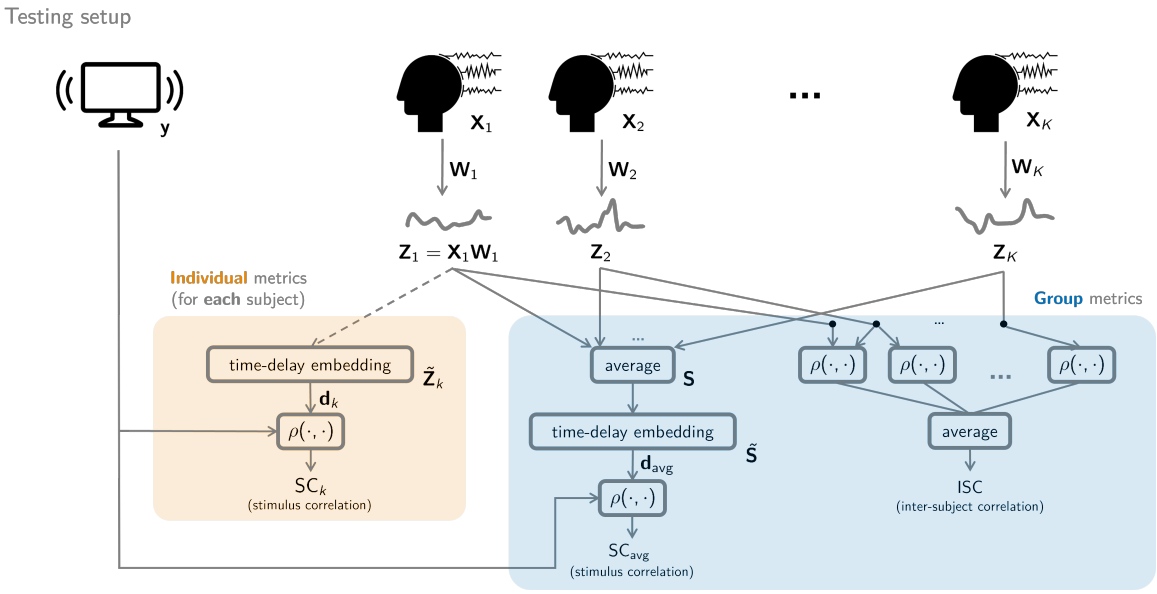

Fig. 2: To compare the different methods, we use three evaluation metrics: one individual metric that quantifies the synchrony between each individual GCCA-decoded signal and the stimulus (i.e., the individual stimulus correlation SC_k per subject k), and two group metrics that quantify the synchrony between the group summary GCCA-decoded signal and the stimulus (i.e., the stimulus correlation SC_{avg} on the average subspace signal) and the average synchrony between GCCA-decoded EEG signals (i.e., the inter-subject correlation (ISC)).

3) Significance level computation: The significance levels of the ISC and SC are computed using a random permutation test, where all correlation between the data is removed by randomly permuting all trials of the different subjects w.r.t. each other or randomly permuting the (projected) EEG and stimulus trials w.r.t. each other. Firstly, 50 Monte Carlo runs are performed in which the 40 training trials are randomly selected and the GCCA and stimulus decoders are trained. At test time, for each Monte Carlo run, 20 random permutations of the test trials are conducted, leading to a total of 1000 resamplings to determine the null distribution of the ISC and SC (i.e., when all correlated patterns are removed). From these null distributions, the 5%-significance level can then be computed and is the same for every algorithm. This significance level of the correlations is mainly determined by the window length ($60\,\text{s}$) and the number of subjects/group size (which is varied from 2 to the maximal number of subjects in the dataset) used to compare the correlation.

D. Decoder setup

1) Filter design: In both the speech and video case, the neural decoders are modeled using a spatiotemporal filter that linearly combines the different EEG channels on different time lags (see Section II-A). The number of time lags L is chosen equal to 5, corresponding to an integration window of $[-250, 250]\text{ms}$ for speech, whereas in the video case, this corresponds to $[-66.7, 66.7]\text{ms}$ (in accordance with [13]).

The additional stimulus decoder to compute the SC in Section IV-C.2.b is similarly modeled as a spatio-temporal decoder, but where all time lags are now chosen post-stimulus, i.e., from 0 to $L_d - 1$ *after* the current stimulus sample. In both cases, the integration window is chosen equal to $[0, 250]\text{ms}$ (post-stimulus), corresponding to $L_d = 3$ in the speech case and $L_d = 9$ in the video case, consistent with, e.g., [3], [14].

In the SI-GCCA/-corrCA estimation, the stimulus representation is augmented with additional time-lagged copies to also allow for temporal filtering (via $\mathbf{V}$) at the stimulus side (including compensation for the intrinsic delay between the stimulus and EEG response), resulting in a Hankel matrix $\mathbf{Y}$. In the speech case, the integration window is chosen equal to $[-1.25, 0]\text{s}$, i.e., preceding the current sample (and response), resulting in a Hankel matrix with $P = 11$ columns. In the video case, the integration window is chosen consistent with [13] equal to $[-500, 0]\text{ms}$, resulting in a Hankel matrix with $P = 16$ columns.

In the speech case, maximally $Q = 32$ components are extracted, whereas this is $Q = 10$ in the video case.

2) Hyperparameter selection: The hyperparameter μ in the GCCA (1) and corrCA (4) estimation problems, determining the weight on the diagonal loading regularization of the neural decoders, and the hyperparameter γ in the SI-GCCA (6) and SI-corrCA (14) estimation problems, determining the weight on the stimulus part, are selected based on the average ISC on a validation set, independent from the training and test set. The optimal hyperparameter is selected based on the maximal average ISC for the first component across the $1\,\text{min}$-trials in the validation set. For μ, a sweep in the range of $\{0, 10^{-5}, 10^{-4.5}, \ldots, 10^{5}\}$ is performed. For γ a sweep in the range of $\{0, 10^{-2}, 10^{-1.5}, \ldots, 10^{8}\}$ is performed.

To not further complicate the hyperparameter search in the speech case, in the SI-GCCA (6) and SI-corrCA (14) estimation problems, the hyperparameter μ for diagonal loading is not validated, but automatically (heuristically) determined using the method suggested by Ledoit and Wolf [48]. In the video case, the optimal validated μ from the GCCA problem is used in the SI-GCCA problem, as the Ledoit-Wolf procedure did not lead to satisfying results in this case.

V. Results and Discussion

We compare GCCA/corrCA with and without diagonal loading from Section II with the newly proposed stimulus-informed counterparts from Section III both on the speech and video dataset according to the experiment details from Section IV. First, we consider only the speech dataset, varying three variables: the amount of available training data (Section V-A), the group size (Section V-B), and the number of EEG channels (Section V-C). In Section V-D, we investigate how the interaction of these three variables can influence the comparison. While the former experiments are driven from the perspective of using the stimulus-informed version to help the estimation of the neural decoders, in Section V-E, we take the alternative perspective of using SI-GCCA to steer the estimation of components towards the chosen stimulus representation.

In Section V-F, we then investigate and explain the performance on the video dataset.

A. Amount of training data

In this experiment, we vary the amount of available training data on the speech dataset from $1\,\mathrm{min}$ to $50\,\mathrm{min}$ while keeping the other variables fixed, as explained in Section IV-C. Per amount of training data, 50 Monte Carlo runs of randomly picking the training trials are performed. Investigating different (smaller) amounts of training data is especially relevant, for example, in a time-adaptive context, when the decoder has to be updated regularly to cope with the non-stationarities in the data [4]. Figure 3a-i shows the ISC as a function of the amount of training data only for the first component, whereas Figure 3a-ii shows the ISC per component when using $15\,\mathrm{min}$ of training data. Figure 3a-iii shows the $\mathrm{SC}_{\mathrm{avg}}$ of the average decoded subspace $\mathbf{S}$ as a function of the amount of training data when using all $Q = 32$ components. The no regularization ('noReg') case refers to $\mu = 0$ in (1) (for GCCA) and (4) (for corrCA), whereas the regularized case ('reg') refers to the case with μ selected based on the validation set performance.

A first important observation from Figure 3a-i, and from Figure 3 in general is that significant ISCs can be obtained using data-driven group decoding algorithms. This shows that synchronized stimulus-following responses across subjects exist and can be decoded. The latter has already been established in other studies on neural responses to natural stimuli [5], [6], [8], and the ISC values in Figure 3 are in line with those observed in other studies. In general, ISC and SC values are typically low because of the very low SNR of the stimulus-following neural responses w.r.t. the background EEG activity.

From Figure 3a-i, we learn that SI-GCCA outperforms GCCA, especially when a smaller amount of training data is available. Firstly, it can be seen that traditional GCCA without any regularization or side-information heavily suffers from overfitting in this case: only when more than $35\,\mathrm{min}$ of training data are available, a significant ISC is found. With diagonal loading, this overfitting effect can be counteracted by using the prior information that norms of the neural decoders should be limited, leading to a significant ISC already with $5\,\mathrm{min}$ of training data. However, a smarter way of introducing side information seems to be using the stimulus as proposed in the SI-GCCA algorithm: especially for smaller amounts of training data, this outperforms GCCA, leading to significant ISCs for $3\,\mathrm{min}$ of training data. Note that from a numerical perspective, the ℓ_2-norm regularization and stimulus information are different: in the former, no additional parameters need to be estimated when using this prior information, whereas in SI-GCCA, additional parameters are introduced in the problem. When increasing the amount of training data, the difference between SI-GCCA and GCCA versions becomes smaller as the additional side-information introduced by the stimulus is outweighed by the large amount of training data, effectively compensating for the high dimensionality of the problem.

Another effective way of coping with the smaller amount of training data is by drastically decreasing the dimensionality of the estimation problem as in the (SI-)corrCA problems. However, there seems to be no additional benefit from the SI-corrCA method w.r.t. corrCA with diagonal loading for extremely low amounts of training data. The corrCA regularization technique proves to be very effective for very low amounts of training data: below $10\,\mathrm{min}$, it outperforms SI-GCCA. However, the flip side is that the performance quickly saturates when the amount of training data increases. (SI-)corrCA clearly suffers from its inability to model subject differences w.r.t. (SI-)GCCA, leading to vastly lower ISCs.

Figure 3a-ii additionally shows that SI-GCCA mainly boosts the ISC for the most significant components, whereas in (SI-)corrCA, more significant components can be found - the ISC is more spread out. In principle, the components cannot be compared between methods one-to-one, as they might represent different activities and only jointly form the basis for the subspace. Therefore, Figure 3a-iii makes it easier to compare across different components, as here the average subspaces across methods are compared in terms of their $\mathrm{SC}_{\mathrm{avg}}$. As expected because of the design of the SI-GCCA method, it is more effective as a preprocessing tool for stimulus decoding, yielding higher stimulus correlations.

B. Group size

In this experiment, we vary the group size from 2 to 19 on the speech dataset while keeping the other variables fixed. Similarly as before, 50 Monte Carlo runs of randomly picking the training trials *and* the group of subjects for a specific group size are performed. This experiment emulates situations where the group size is limited.

Figure 3b-i shows an even clearer effect of the stimulus information w.r.t. the uninformed counterparts when the group size is limited, not only for SI-GCCA but now also for SI-corrCA. The stimulus side-information effectively compensates when less information is available due to a smaller group size, leading to a significant[5] ISC already for 3 subjects in the case of SI-GCCA and only a minor decrease in ISC. Moreover, the corrCA variants are now outperformed each

[5]Note that the significance level decreases with an increasing group size, as when more subjects are available, more pairwise correlations are averaged in the ISC (16), suppressing more potential spurious correlations.

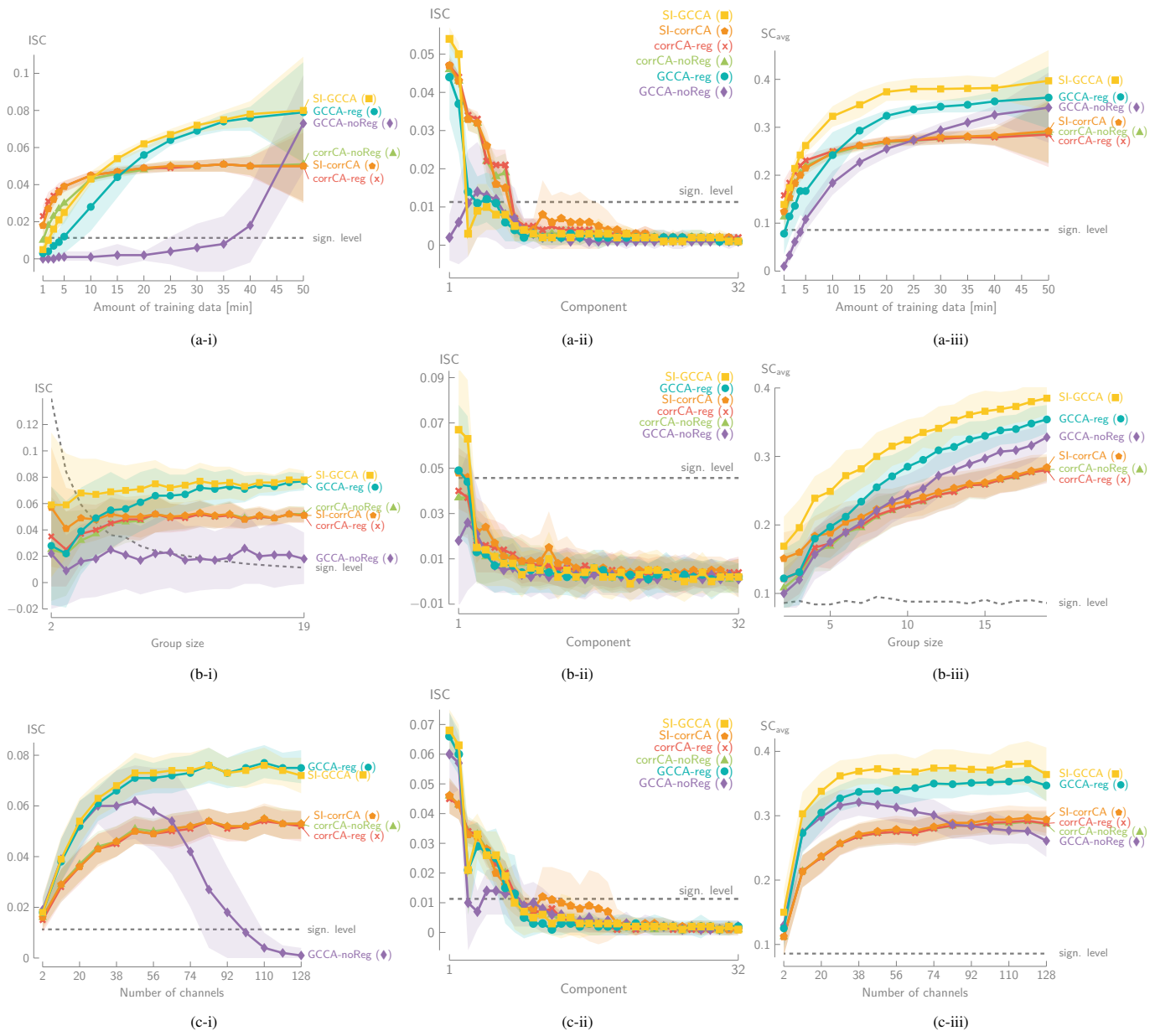

Fig. 3: (-i) The ISC as a function of the amount of training data (a-i), group size (b-i), and number of channels (c-i) for the first component on the speech data, when using 64 channels and all 19 subjects (mean $\pm$ standard deviation across runs). (a-i) For smaller amounts of training data, the SI-GCCA algorithm outperforms the GCCA algorithms, whereas the corrCA versions perform best for extremely low amounts of training data. (b-i) The stimulus-informed versions outperform the uninformed traditional versions for smaller group sizes. (c-i) GCCA clearly overfits without any regularization. (-ii) The ISC across components, for 15 min of training data (a-ii), group size 5 (b-ii), and 38 (random) channels (c-ii), showing mainly an effect of the inclusion of the stimulus on the most significant components. (-iii) The SC_{avg} of the average subspace with $Q = 32$ components as a function of the amount of training data (a-iii), group size (b-iii), and number of channels (c-iii), showing how the stimulus information leads to a better preprocessor (a-iii and b-iii) and again how GCCA without any regularization overfits when using many EEG channels (c-iii).

time by a GCCA counterpart, as sufficient (40 min) training data is available (see Figure 3a-i).

In Figure 3b-ii, it can again be seen that the stimulus information mainly boosts the most significant components, while SI-GCCA/SI-corrCA are also more effective preprocessing tools w.r.t. the stimulus-uninformed traditional versions, as seen in Figure 3b-iii.

C. Number of channels

In this experiment, we vary the number of channels from 2 to 128 in steps of 9 on the speech dataset. Similarly as before, 50 Monte Carlo runs of randomly picking the training trials *and* the EEG channels (randomly chosen) for a specific number of channels are performed. As such, we vary the number of parameters to be estimated *and* the available information.

Figure 3c-i and 3c-iii very clearly show the effect of overfitting on the GCCA method when the number of channels increases, when not using any regularization method or side-information. However, there seems to be no difference between the stimulus-informed and -uninformed (but regularized with diagonal loading) versions in Figure 3c-i and 3c-ii when varying the number of channels. While SI-GCCA results in

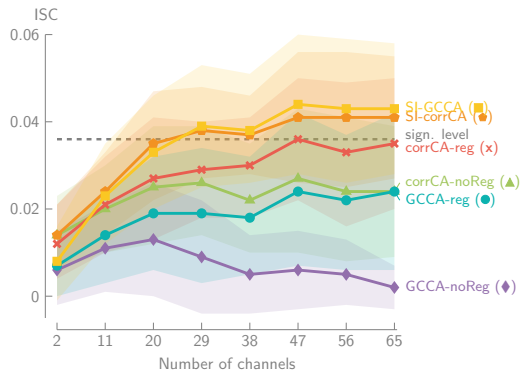

Fig. 4: Both SI-GCCA and SI-corrCA clearly outperform their uninformed counterparts across various numbers of channels when only 15 min of training data and 6 subjects are available (first component).

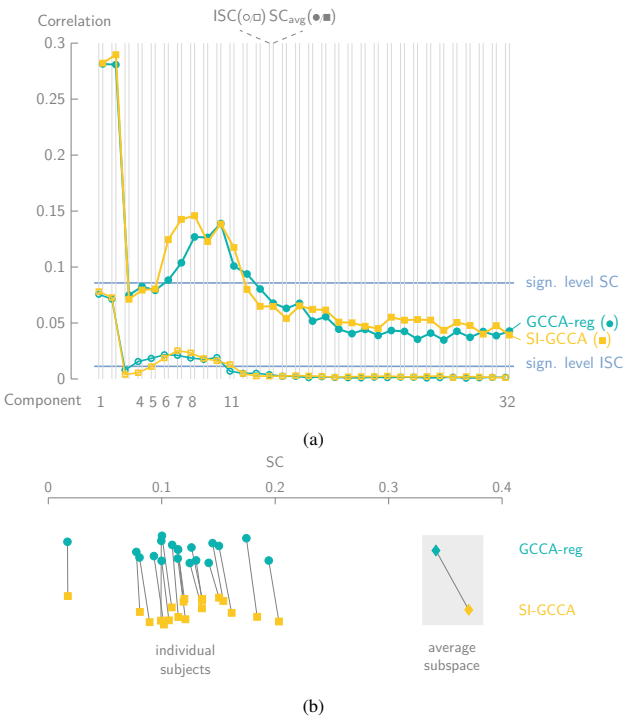

Fig. 5: (a) The ISC between SI-GCCA and GCCA with diagonal loading is very similar across components when using 40 min of training data, all 19 subjects, and the default 64 channels. Mostly for components 6 to 8, SI-GCCA leads to a higher SC_{avg} than GCCA. (b) For all individual subjects, the SC when using the first 11 components extracted with (SI-)GCCA is higher for SI-GCCA than GCCA (Wilcoxon signed-rank test: $n = 19$, p-value < 0.001). This effect is even more amplified when working with the 11-dimensional average subspace.

a slight improvement w.r.t. GCCA for the SC (Figure 3c-iii), the stimulus seems to not help in this situation in terms of the ISC, especially when the other variables (amount of training data and group size) are equal to the (relatively large) default values. However, in Section V-D, we investigate what happens when these three variables interact, and we will show that the number of channels can have an influence when less ideal variable values are used.

D. Interaction of variables

In the previous experiments, only one variable is varied each time, while the other variables are taken constant at their default values as explained in Section IV-C. As such, the previous results are, in a certain sense, still quite conservative, as two of the three variables are each time taken to be quite ideal in the sense that they already lead to data-rich settings, even when the third variable is set to a low value. In this section, we explore how these variables interact and influence the comparison by taking less ideal values for all three variables, which can easily occur in a practical application. More specifically, as a representative example, we choose the amount of training data equal to 15 min, the group size equal to 6, and vary the number of channels from 2 to 65 in steps of 9 on the speech dataset.

Figure 4 shows how this interaction between variables favors even more the stimulus-informed versions w.r.t. the stimulus-uninformed GCCA variants across multiple numbers of channels. Only when using the stimulus, a significant ISC can be obtained when using more than 30 channels. Moreover, similarly to small group sizes in Figure 3b-i, SI-corrCA substantially outperforms corrCA with diagonal loading. This particular instance showcases how in non-ideal, practical use cases, where the amount of training data *and* the group size is limited, our newly proposed SI-GCCA/corrCA algorithms can lead to a substantial benefit.

E. Steering the GCCA estimation

In Sections V-A to V-D, we have shown how SI-GCCA is superior when the available information to estimate the correlated components is limited, e.g., because the amount of training data or group size is limited. Interpreted broadly (not numerically), SI-GCCA can be seen as a task-informed regularization technique that allows to introduce additional available information when estimating correlated components from stimulus-synchronized EEG activity. In this section, we want to put forward an alternative interpretation of SI-GCCA, i.e., using the stimulus to steer the estimation of the correlated components towards the stimulus. This alternative perspective connects with employing (SI-)GCCA, e.g., as a preprocessor for stimulus decoding [26] or to use the ISC as a proxy for attentional engagement to the content of one particular stimulus, for example, the teacher's voice in a classroom [5].

From Figure 3c-i and 3c-iii, it can already be seen that for a similar ISC, SI-GCCA can still lead to higher SC w.r.t. GCCA. This already indicates that while the extracted components per subject are almost equally well correlated with one another, the ones extracted by SI-GCCA are still more related to the stimulus (feature) than for GCCA, in other words, SI-GCCA steers these components more towards the stimulus. Whereas Figure 3c-iii shows only the SC for the average subspace, we further investigate whether this steering behavior is also present on an individual level, per component and subject. Therefore, we more closely compare SI-GCCA and GCCA with diagonal loading for 40 min of training data, using all 19 subjects and the default 64 EEG channels.

As could be seen already in Figure 3a-i for the default settings and is now confirmed in Figure 5a, especially for the most significant components, there is hardly any difference in

terms of ISC. When summing the ISCs across all components, GCCA and SI-GCCA lead to the same cumulative ISC of 0.335, and no significant difference can be found across components (Wilcoxon signed rank test, $n = 32$, p-value$= 0.24$). However, when we have a look at the SC per individual component in Figure 5a, SI-GCCA almost always leads to a higher or comparable SC than GCCA, even when the ISC is lower (e.g., for component 4, 5, 6). This is specifically noticeable for components 6 to 8. Moreover, in Figure 5b, we show the SC when using the 11-dimensional subspace (all components beyond the 11th are not significant in Figure 5a) from (SI-)GCCA, per individual subject and also for the average subspace. For all subjects, the SC is higher for the SI-GCCA method compared to the GCCA method, showing a significant improvement across subjects (Wilcoxon signed-rank test: $n = 19$, p-value < 0.001). Furthermore, this effect is amplified when using the average subspace. This shows how SI-GCCA can also be used to steer the correlated components to be more correlated with the stimulus and shows its benefit as a preprocessor to boost the SNR before stimulus decoding, both on an individual and group level.

F. Analysis of video data with object-based optical flow and the effect of the specific stimulus representation

In this section, we evaluate and compare SI-GCCA with GCCA on the video dataset.

In Figure 6a, the ISC is shown as a function of group size when using $40\,\text{min}$ of training data and all 64 EEG channels. From Figure 3b-i, we would expect from the speech dataset that the stimulus information would especially help for group sizes below 10. However, Figure 6a shows hardly any difference between SI-GCCA and GCCA, indicating that the stimulus does not improve the estimation of the correlated components (but also does not deteriorate it). An explanation can be found in Yao et al. [13], where it is shown that the object-based optical flow only explains 6.9% of the variance present in the correlated components across the EEG's of the different subjects. A large proportion of the variance in the stimulus-related neural responses is thus not yet explained, such that using this specific video feature in SI-GCCA has only a minor effect on the estimated correlated components in terms of maximizing the ISC. To strengthen this conclusion, we have performed a similar experiment as in [13] to quantify the variance explained in the GCCA components for the speech dataset when using the speech envelope as a feature. When regressing out the speech envelope per subject and re-estimating GCCA (more details in [13]), we find that the speech envelope explains around 40.6% of all correlated activity, explaining why SI-GCCA performs much better on the speech dataset. This observation entails an important limitation when using our proposed SI-GCCA method: its impact is bound to the specific chosen stimulus representation(s).

However, even when the stimulus representation does not explain much of the correlated neural (stimulus-related) activity as is the case here for object-based optical flow in the video dataset, SI-GCCA can still be used to steer the correlated components towards that specific stimulus representation. We highlight this point by comparing the ISC on

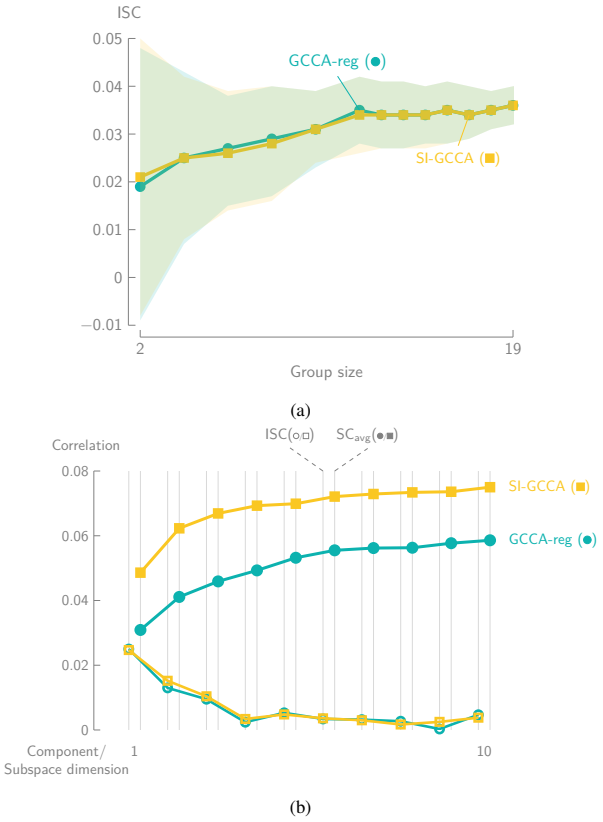

Fig. 6: (a) The ISC as a function of group size with $40\,\text{min}$ of training data and 64 EEG channels for the video dataset. There is no benefit from the stimulus-informed version. (b) There is hardly any difference when comparing the ISC between SI-GCCA and GCCA across components for group size 4. However, SI-GCCA clearly leads to a higher SC_{avg} using increasing subspace dimension, showing its capability to steer the estimation of the correlated components towards the stimulus representation.

the video dataset for group size 4 with the SC_{avg} across all components. While Figure 6b shows hardly any difference in ISC per individual correlated components, the subspaces of increasing dimension are clearly much more related to the object flow feature, already starting from a one-dimensional subspace. This improved SC does not reduce the correlation across subjects, as the ISC remains the same between both methods. This shows that SI-GCCA extracts equally correlated components as GCCA (across subjects), but the former extracts components that better capture the temporal dynamics in the optical flow related to the moving object in the video.

VI. Conclusion

In this paper, we proposed a new algorithm for the group analysis of stimulus-following neural responses within a group of people attending to the same natural stimulus. Our proposed framework allows to take the stimulus itself into account when estimating the correlated components across subjects with GCCA and its subject-generic variant corrCA. This stimulus-informed GCCA framework can still be solved as a GEVD, inheriting this attractive property from MAXVAR-GCCA.

We compared our newly proposed SI-GCCA algorithms with the traditional stimulus-uninformed versions on a speech and video dataset, using the speech envelope and object-based optical flow as exemplary stimulus representations. We

demonstrated the superiority of using the stimulus as side-information when the amount of training data or group size is limited, even more so when these different variables interact, also with the number of channels. This shows its practical relevance, for example, in situations where the training set size is limited, e.g., in a context of time-adaptive, online processing, or when the group size is limited, e.g., as determined by the application. Using the video dataset, we showed that a limitation of SI-GCCA, besides the requirement of having access to the stimulus, is its dependency on the specific stimulus representation.

Besides using SI-GCCA to introduce the stimulus as valuable side-information to robustify the estimation of correlated components across a group of subjects when the available estimation data is a priori limited, it can also be used to steer the correlated components explicitly in the direction of the stimulus (features). To this end, we showed that a higher stimulus correlation can be obtained when using SI-GCCA versus the traditional uninformed GCCA, without any significant reduction in the ISC.

To sum up, the proposed SI-GCCA algorithm can be employed for various purposes in the group decoding of stimulus-following neural responses, e.g., to avoid overfitting and compensate for limited available information, or to steer the design of the neural decoders towards a specific stimulus representation. As such, it can enable various applications, ranging from more fundamentally-oriented (e.g., preprocessing and dimensionality reduction) to application-specific (e.g., quantifying attention in the classroom).

[1] L. F. Nicolas-Alonso and J. Gomez-Gil, "Brain Computer Interfaces, a Review," *Sensors*, vol. 12, no. 2, pp. 1211–1279, 2012.

[2] F. Lotte, L. Bougrain, A. Cichocki, M. Clerc, M. Congedo, A. Rakotomamonjy, and F. Yger, "A review of classification algorithms for EEG-based brain–computer interfaces: a 10 year update," *Journal of Neural Engineering*, vol. 15, no. 3, p. 031005, 2018.

[3] S. Geirnaert, S. Vandecappelle, E. Alickovic, A. de Cheveigné, E. C. Lalor, B. T. Meyer, S. Miran, T. Francart, and A. Bertrand, "Electroencephalography-Based Auditory Attention Decoding: Toward Neurosteered Hearing Devices," *IEEE Signal Processing Magazine*, vol. 38, no. 4, pp. 89–102, 2021.

[4] S. Geirnaert, T. Francart, and A. Bertrand, "Time-adaptive Unsupervised Auditory Attention Decoding Using EEG-based Stimulus Reconstruction," *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 8, pp. 3767–3778, 2022.

[5] A. T. Poulsen, S. Kamronn, J. P. Dmochowski, L. C. Parra, and L. K. Hansen, "EEG in the classroom: Synchronised neural recordings during video presentation," *Scientific Reports*, vol. 7, pp. 1–9, 2017.

[6] J. P. Dmochowski, P. Sajda, J. Dias, and L. C. Parra, "Correlated components of ongoing EEG point to emotionally laden attention - a possible marker of engagement?" *Frontiers in Human Neuroscience*, vol. 6, p. 112, 2012.

[7] J. P. Dmochowski, M. A. Bezdek, B. P. Abelson, J. S. Johnson, E. H. Schumacher, and L. C. Parra, "Audience preferences are predicted by temporal reliability of neural processing," *Nature Communications*, vol. 5, no. 1, pp. 1–9, 2014.

[8] J. J. Ki, S. P. Kelly, and L. C. Parra, "Attention strongly modulates reliability of neural responses to naturalistic narrative stimuli," *Journal of Neuroscience*, vol. 36, no. 10, pp. 3092–3101, 2016.

[9] J. P. Dmochowski, J. J. Ki, P. DeGuzman, P. Sajda, and L. C. Parra, "Extracting multidimensional stimulus-response correlations using hybrid encoding-decoding of neural activity," *NeuroImage*, vol. 180, pp. 134–146, 2018.

[10] J. Vanthornhout, L. Decruy, and T. Francart, "Effect of Task and Attention on Neural Tracking of Speech," *Frontiers in Neuroscience*, vol. 13, p. 977, 2019.

[11] J. Belo, M. Clerc, and D. Schön, "EEG-Based Auditory Attention Detection and Its Possible Future Applications for Passive BCI," *Frontiers in Computer Science*, vol. 3, no. 661178, 2021.

[12] M. Nentwich, M. Leszczynski, B. E. Russ, L. Hirsch, N. Markowitz, K. Sapru, C. E. Schroeder, A. D. Mehta, S. Bickel, and L. C. Parra, "Semantic novelty modulates neural responses to visual change across the human brain," *Nature Communications*, vol. 14, no. 1, p. 2910, 2023.

[13] Y. Yao, A. Stebner, T. Tuytelaars, S. Geirnaert, and A. Bertrand, "Identifying Temporal Correlations Between Natural Single-shot Videos and EEG Signals," *Journal of Neural Engineering*, vol. 21, no. 1, p. 016018, 2024.

[14] J. A. O'Sullivan, A. J. Power, N. Mesgarani, S. Rajaram, J. J. Foxe, B. G. Shinn-Cunningham, M. Slaney, S. A. Shamma, and E. C. Lalor, "Attentional Selection in a Cocktail Party Environment Can Be Decoded from Single-Trial EEG," *Cerebral Cortex*, vol. 25, no. 7, pp. 1697–1706, 2014.

[15] S. Geirnaert, T. Francart, and A. Bertrand, "Unsupervised Self-Adaptive Auditory Attention Decoding," *IEEE Journal on Biomedical and Health Informatics*, vol. 25, no. 10, pp. 3955–3966, 2021.

[16] J. Vanthornhout, L. Decruy, J. Wouters, J. Z. Simon, and T. Francart, "Speech Intelligibility Predicted from Neural Entrainment of the Speech Envelope," *Journal of the Association for Research in Otolaryngology*, vol. 19, no. 2, pp. 181–191, 2018.

[17] I. Davidesco, C. Matuk, D. Bevilacqua, D. Poeppel, and S. Dikker, "Neuroscience Research in the Classroom: Portable Brain Technologies in Education Research," *Educational Researcher*, vol. 50, no. 9, pp. 649–656, 2021.

[18] V. Delvigne, H. Wannous, J.-P. Vandeborre, L. Ris, and T. Dutoit, "Attention Estimation in Virtual Reality with EEG based Image Regression," in *2020 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)*, 2020, pp. 10–16.

[19] A. Roebben, N. Heintz, S. Geirnaert, T. Francart, and A. Bertrand, "'Are you even listening?' - EEG-based detection of absolute auditory attention to natural speech," *Journal of Neural Engineering*, 2024.

[20] S. J. Aiken and T. W. Picton, "Human Cortical Responses to the Speech Envelope," *Ear and Hearing*, vol. 29, no. 2, pp. 139–157, 2008.

[21] N. Ding and J. Z. Simon, "Neural coding of continuous speech in auditory cortex during monaural and dichotic listening," *Journal of Neurophysiology*, vol. 107, no. 1, pp. 78–89, 2012.

[22] N. Mesgarani and E. F. Chang, "Selective cortical representation of attended speaker in multi-talker speech perception," *Nature*, vol. 485, pp. 233–236, 2012.

[23] D. D. Wong, S. A. Fuglsang, J. Hjortkjær, E. Ceolini, M. Slaney, and A. de Cheveigné, "A Comparison of Regularization Methods in Forward and Backward Models for Auditory Attention Decoding," *Frontiers in Neuroscience*, vol. 12, no. 531, 2018.

[24] S. Miran, A. Presacco, J. Z. Simon, M. C. Fu, S. I. Marcus, and B. Babadi, "Dynamic estimation of auditory temporal response functions via state-space models with Gaussian mixture process noise," *PLoS Computational Biology*, vol. 16, no. 8, p. e1008172, 2020.

[25] A. de Cheveigné, G. M. Di Liberto, D. Arzounian, D. D. Wong, J. Hjortkjær, S. Fuglsang, and L. C. Parra, "Multiway canonical correlation analysis of brain data," *NeuroImage*, vol. 186, no. October 2018, pp. 728–740, 2019.

[26] G. M. Di Liberto, G. Marion, and S. A. Shamma, "Accurate Decoding of Imagined and Heard Melodies," *Frontiers in Neuroscience*, vol. 15, no. 673401, 2021.

[27] J. D. Carroll, "Generalization of canonical correlation analysis to three of more sets of variables," in *Proceedings of the 76th annual convention of the American Psychological Association*, vol. 3, 1968, pp. 227–228.

[28] J. R. Kettenring, "Canonical analysis of several sets of variables," *Biometrika*, vol. 58, no. 3, pp. 433–451, 1971.

[29] H. Hotelling, "Relations Between Two Sets of Variates," *Biometrika*, vol. 28, pp. 321–377, 1936.

[30] L. Chen, P. Chen, S. Zhao, Z. Luo, W. Chen, Y. Pei, H. Zhao, J. Jiang, M. Xu, Y. Yan, and E. Yin, "Adaptive asynchronous control system of robotic arm based on augmented reality-assisted brain–computer interface," *Journal of Neural Engineering*, vol. 18, no. 6, p. 066005, 2021.

[31] A. de Cheveigné, D. D. Wong, G. M. Di Liberto, J. Hjortkjær, M. Slaney, and E. C. Lalor, "Decoding the auditory brain with canonical component analysis," *NeuroImage*, vol. 172, pp. 206–216, 2018.

[32] Y. Zhang, G. Zhou, J. Jin, X. Wang, and A. Cichocki, "Frequency recognition in SSVEP-based BCI using multiset canonical correlation analysis," *International Journal of Neural Systems*, vol. 24, no. 04, p. 1450013, 2014.

[33] S. Geirnaert, T. Francart, and A. Bertrand, "Stimulus-Informed Generalized Canonical Correlation Analysis of Stimulus-Following Brain Responses," in *2023 31st European Signal Processing Conference (EUSIPCO)*, 2023, pp. 1000–1004.

[34] M. Sørensen, C. I. Kanatsoulis, and N. D. Sidiropoulos, "Generalized Canonical Correlation Analysis: A Subspace Intersection Approach," *IEEE Transactions on Signal Processing*, vol. 69, pp. 2452–2467, 2021.

[35] X. Fu, K. Huang, E. E. Papalexakis, H.-A. Song, P. P. Talukdar, N. D. Sidiropoulos, C. Faloutsos, and T. Mitchell, "Efficient and distributed algorithms for large-scale generalized canonical correlations analysis," in *2016 IEEE 16th International Conference on Data Mining (ICDM)*, 2016, pp. 871–876.

[36] P. Horst, *Generalized canonical correlations and their application to experimental data*. Journal of Clinical Psychology, 1961, no. 14.

[37] C. Hovine and A. Bertrand, "Distributed MAXVAR: Identifying Common Signal Components across the Nodes of a Sensor Network," in *2021 29th European Signal Processing Conference (EUSIPCO)*. EURASIP, 2021, pp. 2159–2163.

[38] A. Hassani, A. Bertrand, and M. Moonen, "GEVD-Based Low-Rank Approximation for Distributed Adaptive Node-Specific Signal Estimation in Wireless Sensor Networks," *IEEE Transactions on Signal Processing*, vol. 64, no. 10, pp. 2557–2572, 2016.

[39] J. Vía, I. Santamaría, and J. Pérez, "A learning algorithm for adaptive canonical correlation analysis of several data sets," *Neural Networks*, vol. 20, no. 1, pp. 139–152, 2007.

[40] S. Geirnaert, Y. Yao, T. Francart, and A. Bertrand, "SI-GCCA MATLAB Toolbox and Experiments," 2024. [Online]. Available: https://github.com/AlexanderBertrandLab/si-gcca

[41] M. P. Broderick, A. J. Anderson, G. M. Di Liberto, M. J. Crosse, and E. C. Lalor, "Electrophysiological Correlates of Semantic Dissimilarity Reflect the Comprehension of Natural, Narrative Speech," *Current Biology*, vol. 28, no. 5, pp. 803–809.e3, 2018.

[42] M. P. Broderick, A. J. Anderson, G. M. Di Liberto, M. J. Crosse, and E. C. Lalor, "Data from: Electrophysiological correlates of semantic dissimilarity reflect the comprehension of natural, narrative speech," Feb. 2019. [Online]. Available: https://doi.org/10.5061/dryad.070jc

[43] Y. Yao, A. Stebner, T. Tuytelaars, S. Geirnaert, and A. Bertrand, "Video-EEG Encoding-Decoding Dataset KU Leuven," Jan. 2024. [Online]. Available: https://zenodo.org/doi/10.5281/zenodo.10512413

[44] G. M. Di Liberto, J. A. O'Sullivan, and E. C. Lalor, "Low-Frequency Cortical Entrainment to Speech Reflects Phoneme-Level Processing," *Current Biology*, vol. 25, no. 19, pp. 2457–2465, 2015.

[45] C. Brodbeck, L. E. Hong, and J. Z. Simon, "Rapid Transformation from Auditory to Linguistic Representations of Continuous Speech," *Current Biology*, vol. 28, no. 24, pp. 3976–3983.e5, 2018.

[46] M. Gillis, J. Vanthornhout, J. Z. Simon, T. Francart, and C. Brodbeck, "Neural Markers of Speech Comprehension: Measuring EEG Tracking of Linguistic Speech Representations, Controlling the Speech Acoustics," *Journal of Neuroscience*, vol. 41, no. 50, pp. 10 316–10 329, 2021.

[47] X. Zhang, J. Li, Z. Li, B. Hong, T. Diao, X. Ma, G. Nolte, A. K. Engel, and D. Zhang, "Leading and following: Noise differently affects semantic and acoustic processing during naturalistic speech comprehension," *NeuroImage*, vol. 282, p. 120404, 2023.

[48] O. Ledoit and M. Wolf, "A well-conditioned estimator for large-dimensional covariance matrices," *Journal of Multivariate Analysis*, vol. 88, no. 2, pp. 365–411, 2004.