Published in Biomedical Signal Processing and Control, vol. 61, article no. 101948, 2020. DOI: 10.1016/j.bspc.2020.101948

Joint-IVA for Identification of Discriminating Features in EEG: Application to a Driving Study

Ben Gabrielson*, M. A. B. S. Akhonda, Suchita Bhinge,
Justin Brooks, Qunfang Long, and Tülay Adali*

***Abstract*—We propose a new method, joint independent vector analysis (jIVA), for obtaining *discriminating features*, i.e., interpretable signatures from medical data that can be used to study differences between multiple conditions or groups. The method is especially attractive for event related studies of electroencephalogram (EEG) data as it enables one to take advantage of the cross information across multiple channels effectively while enabling the use of information from multiple epochs. We introduce the general model and then demonstrate its successful application to EEG data collected during a driving experiment. As opposed to traditional analysis techniques that only detect differences, we identify statistically significant differences in measured band power showing when and how the differences occur for two experimental conditions across the same group of subjects. We compare jIVA features to those produced from competing data-driven approaches and demonstrate the advantages of jIVA as it fully leverages the statistical dependencies across multiple electrodes, and note its promise as a powerful data-driven method of obtaining informative features of multiset data.**

I. INTRODUCTION

Data driven techniques make only a few assumptions about the nature of the data, and can produce informative features that can be used for multiple purposes. Blind source separation (BSS) techniques are an important class among those, and in particular independent component analysis (ICA) has demonstrated much success in identifying informative features describing patterns within various types of datasets, e.g. [1], [2], [3], [4]. These techniques are often useful when one may reasonably assume that an observed dataset can be described by an associated generative model, typically a linear mixture model. In the context of datasets obtained from electroencephalogram (EEG) measurements, ICA and more recently developed independent vector analysis (IVA)—a multiset extension of ICA—have both shown success in obtaining informative features representing task and non-task related neurological changes, as well as identifying features representing various artifacts [4], [5], [6], [7], [8].

As ICA and IVA have become increasingly popular for analyzing single and multiple datasets, the strong desire to leverage shared information between different datasets (i.e. linking multiple sets of observations or multiple measurement periods) has led to the development of multiple ICA- and IVA-based data fusion models [9], [10], [11], [12], [13], [14]. These models operate via processing of multiple sets of observations concurrently to generate informative features, which describe shared information or information distinct to each set of observations—where sets of observations may be of the same type, i.e. multi-electrode datasets for EEG data, or of different types, such as EEG, fMRI, among others. We refer to the former group as multi-set and the second as multi-modal. One such successful data fusion method is joint-ICA (jICA), where multiple sets of observations are concatenated horizontally, assuming a common mixing matrix, thus leveraging shared information among the observations to provide decompositions better representative of this shared information, while also potentially yielding more robust ICA decompositions due to large sample properties on concatenated datasets [15]. IVA is another successful method of data fusion, relaxing the common mixing matrix assumption and allowing multiple datasets to interact when providing individual decompositions per each dataset, leveraging shared information (dependence) between these datasets to generate more robust features describing these dependencies [16]. While jICA can be used both for multiset and multimodal datasets directly, IVA is developed for multiset data and needs to be adapted to multimodal data through a transpose operation as described in [11] .

Furthermore, data fusion techniques such as jICA and IVA have shown to be effective ways of obtaining discriminating features: informative features describing differences between different datasets, such as with two or more "conditions" in the observed data (e.g. in EEG or fMRI, eyes open vs. eyes closed) [17], [18], [19]. These techniques are often better than conventional methods as they are fully multivariate and hence can account for the variability in the data, and minimize modeling assumptions. As an example, when discriminating between conditions in EEG datasets, averages of epochs represented by event-related potentials (ERPs) or event related spectral perturbations (ESRPs) across different conditions may result in important variability being averaged out, e.g. if

This work was supported by the U.S. Army Research Laboratory and the Cognition and Neuroergonomics Collaborative Technology Alliance, W911NF-10-D-0002, under Cooperative Agreement Number W911NF-10-2-0022, and the grants NSF-CCF 1618551 and NSF-NCS 1631838. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein. The authors would like to thank Scott Kerick for his assistance in designing the study, and thank Thomas Drake for his role in preprocessing the experimental data.

B. Gabrielson* is with the Department of Computer Science and Electrical Engineering, University of Maryland, Baltimore County, Baltimore, MD 21250, USA (email: bgabriel@umbc.edu).

Mohammad Abu Bak Siddique Akhonda, Suchita Bhinge, Qunfang Long, Thomas Drake, and Tulay Adali* are with the Department of Computer Science and Electrical Engineering, University of Maryland, Baltimore County, Baltimore, MD, 21250, USA.

Justin Brooks is with the U.S. Army Research Laboratory, Human Research and Engineering Directorate, Aberdeen Proving Ground, MD, 21005, USA.

subject responses are not sufficiently synchronized in time for averaging to be reasonably warranted [20]. As another example for EEG, simply reporting differences in statistical values of spectral power (e.g. mean, kurtosis, etc.) as the discriminating features [21] might fail to describe variability across measurement periods, and may present a less intuitive or informative interpretation of the underlying differences.

While jICA has yielded successful results, its common mixing matrix assumption is a strict one. Another option, applying ICA individually to each dataset, does not allow these datasets to "fully interact", i.e. representing or exploiting shared information across datasets through the assumed dependence structure. In addition, when comparing ICA decompositions across different datasets, source alignment across datasets (i.e., pairing sources between two datasets) is an extra step outside of ICA which in itself may be problematic, due to differences in each ICA decomposition. IVA, on the other hand, can account for the dependence structure when estimating informative features, and is able to produce results superior to ICA alone when there is dependence among the datasets. Many ICA and IVA algorithms usually require a large number of samples to provide robust decompositions, thus we propose a solution that leverages the strengths of both jICA and IVA, and introduce a new method joint-IVA (jIVA) for identifying discriminating features in EEG datasets.

Thus, we describe jIVA as an effective method for estimating discriminating EEG features. The method performs data fusion in two ways, leveraging large sample properties and shared information between multiple measurement periods, while also leveraging the underlying dependence structure between datasets. After jIVA is performed, discriminating features are detected by applying a statistical test per each jIVA feature (e.g. a paired two sample t-test). We apply the method to a multi-electrode EEG driving study dataset to study differences between two groups of experimental driving conditions, and summarize the extracted discriminating features via clustering measures. Our jIVA framework provides results that generally agree with comparable jICA and IVA decompositions, thus establishing confidence in jIVA results, while additionally providing improved statistical power for source separation, and additionally providing supplementary information as to how variation exists across the datasets.

Unlike simple statistical tests that only determine the presence of group difference, our method extracts temporal signatures of the data that describe how differences are manifest within measurements, how variations of these differences exist across subject measurements, and where and when these differences occur. From a neuroscientist's interpretation of jIVA features estimated from the driving study data, these jIVA features describe how response differences between conditions may characterize different control strategies between conditions. The jIVA features we extract demonstrate that for subjects with prior knowledge of impending perturbation events, these subjects' respective measurements are characterized by a more prominent trend of decrease in frontal theta band power during a perturbation event, and a more prominent trend of decrease in parietal alpha band power immediately afterwards, compared to subjects without prior knowledge of impending events.

The paper is organized as follows. Section II describes the assumptions and formulation of the ICA, jICA, and IVA models, and demonstrates how discriminating features can be detected. Section III describes the jIVA framework for obtaining discriminating features. Section IV explains how we constructed a suitable dataset structure for the jIVA model, given EEG experimental data applicable to the model. Section V discusses how discriminating features are interpreted through clustering measures and averaging of epochs. Section VI shows results from the implementation of the jIVA framework on the EEG data, in comparison to results from similarly based jICA and IVA-based frameworks. Section VII concludes the paper with closing remarks on the model's usefulness compared to other methods for discriminating between conditions.

II. Methods and Materials

A. Motivating Data-Fusion with EEG Data

Consider EEG data from an event related study, with data measuring cortical electrical activity due to a task-related event (e.g., pressing a button). Such studies are often designed with the goal to study group differences in response to different conditions (e.g., when subjects have eyes open, vs eyes closed). When measurements of different conditions are recorded across the same subjects for the given task, an investigator may seek to generate informative features of the data, particularly features that are useful for discriminating between measurements of different conditions. This is useful e.g. to understand brain function, identify biomarkers of a disease/disorder, among other applications.

Such experiments often repeat the task and corresponding measurements multiple times throughout the experiment, defining an "epoch" as a time period of measurement starting before the task and ending sometime after the task is completed. As information useful for discriminating between subject conditions is present in measurements across multiple epochs, while also present across multiple electrodes, these multiple measurements provide increased statistical power that may be exploited to estimate more robust features: features that are more informative and/or reliable in describing differences between the conditions. Data fusion methods based on matrix factorizations are able to let these multiple measurements fully interact, thus matrix factorizations are desirable implementations for obtaining such features.

Blind source separation (BSS) techniques are well suited for generating such informative features, as they allow fusion of all available information from multiple subjects, epochs, and electrodes. BSS techniques are additionally data driven, placing minimum assumptions on the data. Thus, given EEG measurements across multiple subjects, epochs, and electrodes, using these multiple measurements in a fusion based BSS model can better reveal information shared across the multiple measurements, especially information that is useful for discriminating between different conditions.

B. ICA

ICA is a popular matrix factorization technique for achieving BSS, and has proven especially useful for generating informative features of observed data. The ICA model assumes that each observation in a dataset may be modeled as a linear mixture of underlying factors, called "sources". ICA additionally strives to estimate underlying sources with maximized statistical independence, which is an appropriate assumption for many applications [3].

With $\mathbf{s}(t) = [s_1(t), \ldots, s_N(t)]^\top \in \mathbb{R}^N$ denoting the N underlying sources at some time index t ($t = 1, \ldots, T$), ICA assumes that these sources are mixed by an unknown invertible matrix $\mathbf{A} \in \mathbb{R}^{N \times N}$, to produce observed mixtures $\mathbf{x}(t) = [x_1(t), \ldots, x_N(t)]^\top \in \mathbb{R}^N$, $t = 1, \ldots, T$. Thus, the relationship between the mixtures and the sources is represented as:

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t), \qquad t = 1, \ldots, T. \tag{1}$$

Here, $(.)^\top$ denotes the transpose. Across all observed samples $t = 1, \ldots, T$, the ICA formulation may also be represented by the matrix product:

$$\mathbf{X} = \mathbf{A}\mathbf{S} \tag{2}$$

with $\mathbf{X}$, $\mathbf{S} \in \mathbb{R}^{N \times T}$, $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_N]^\top$ and $\mathbf{S} = [\mathbf{s}_1, \ldots, \mathbf{s}_N]^\top$. To estimate the N underlying sources $\mathbf{S}$, ICA estimates a demixing matrix $\mathbf{W} \in \mathbb{R}^{N \times N}$ that maximizes independence between the source estimates $\mathbf{Y}$, given by $\mathbf{Y} = \mathbf{W}\mathbf{X}$, with $\mathbf{Y} \in \mathbb{R}^{N \times T}$. To achieve such demixing, one common approach is to minimize the mutual information (MI) between the estimated sources [22].

The ICA model is often used to individually analyze a given dataset. There are also a number of ICA models designed to jointly analyze multiple datasets, and one such model is the joint-ICA model. Joint-ICA entails concatenating J datasets horizontally into a single matrix $\overline{\mathbf{X}}$, and then performing ICA on this matrix. Thus, jICA is formulated as

$$\overline{\mathbf{X}} = \overline{\mathbf{A}}\,\overline{\mathbf{S}}\,, \tag{3}$$

where $\overline{\mathbf{X}} = [\mathbf{X}_1, \ldots, \mathbf{X}_J]$, and $\overline{\mathbf{S}} = [\mathbf{S}_1, \ldots, \mathbf{S}_J]$. Here, $\mathbf{S}_j \in \mathbb{R}^{N \times T_j}$ are the source estimates of the jth concatenated dataset $\mathbf{X}_j \in \mathbb{R}^{N \times T_j}$ and $\overline{\mathbf{A}} \in \mathbb{R}^{N \times N}$ is the mixing matrix used for all J concatenated datasets. As it is not required to have the same number of samples for all concatenated datasets $\mathbf{X}_j$, the number of samples of the jth dataset is denoted by T_j. The jICA model assumes that the underlying concatenated sources share a common probability distribution function (PDF), and assumes that these concatenated datasets are similarly mixed, thus sharing $\overline{\mathbf{A}}$ [18].

C. IVA

IVA provides a generalization of ICA to multiple datasets. Unlike individual ICA decompositions per each dataset, IVA allows the datasets to interact, allowing a decomposition that provides an improved estimation of sources exhibiting dependencies across datasets [16], [23]. Given K datasets $\mathbf{X}^{[k]} \in \mathbb{R}^{N \times T}$, where each dataset consists of N observations over T samples, the IVA model is formulated as

$$\mathbf{X}^{[k]} = \mathbf{A}^{[k]}\mathbf{S}^{[k]}, \text{ for } k = 1, \ldots, K. \tag{4}$$

Here, $\mathbf{S}^{[k]} \in \mathbb{R}^{N \times T}$ and $\mathbf{A}^{[k]} \in \mathbb{R}^{N \times N}$ are respectively the estimated sources and mixing matrices for the kth dataset. Like ICA, IVA similarly seeks to estimate sources with maximal statistical independence within a given dataset, while additionally maximizing dependence between sources across datasets, to produce improved estimates of sources dependent across datasets. This is achieved by maximizing independence among the N source component vectors (SCVs), which themselves are formed as collections of dependent sources across datasets, demonstrated by the green SCV matrix defined in Fig. 2.

D. Joint-IVA

As data-fusion can improve analysis beyond that of independently analyzing multiple measurements, here, we propose a method for further modeling dependencies across multiple datasets, compared to other fusion models such as jICA and IVA. Our method, jIVA, is based on a reasonable common mixing assumption across concatenated datasets, seen within the jICA model, while allowing us to simultaneously take advantage of cross-information across datasets with source dependence, seen within the IVA model.

Joint-IVA entails forming K data matrices $\overline{\mathbf{X}}^{[k]}$ (for $k = 1, \ldots, K$), with each matrix composed of J horizontally concatenated datasets. Fig. 1 demonstrates the jIVA construction of each electrode data matrix of J concatenated datasets, and Fig. 2 defines the jIVA model given K electrodes and J epochs of EEG data. Thus, the jIVA model is formulated as

$$\overline{\mathbf{X}}^{[k]} = \overline{\mathbf{A}}^{[k]}\,\overline{\mathbf{S}}^{[k]}, \text{ for } k = 1, \ldots, K, \tag{5}$$

where $\overline{\mathbf{X}}^{[k]} = [\mathbf{X}_1^{[k]}, \ldots, \mathbf{X}_J^{[k]}]$, and $\overline{\mathbf{S}}^{[k]} = [\mathbf{S}_1^{[k]}, \ldots, \mathbf{S}_J^{[k]}]$. Here, $\overline{\mathbf{X}}^{[k]}$ is the kth data matrix, $\overline{\mathbf{S}}^{[k]}$ is the source matrix of the kth data matrix, with $\overline{\mathbf{X}}^{[k]}$, $\overline{\mathbf{S}}^{[k]} \in \mathbb{R}^{N \times T}$, and $\overline{\mathbf{A}}^{[k]}$ is the mixing matrix of the kth data matrix, with $\overline{\mathbf{A}}^{[k]} \in \mathbb{R}^{N \times N}$. Within the kth data matrix, the jth concatenated dataset $\mathbf{X}_j^{[k]}$ is assumed to be a mixture of the corresponding sources $\mathbf{S}_j^{[k]}$, with $\mathbf{X}_j^{[k]}$, $\mathbf{S}_j^{[k]} \in \mathbb{R}^{N \times T_{j,k}}$. The number of samples $T_{j,k}$ may vary across concatenated datasets $\mathbf{X}_j^{[k]}$, so long as the total number of samples across concatenated datasets sums up to the same number T for all data matrices $\overline{\mathbf{X}}^{[k]}$, thus $T = \sum_{j=1}^{J} T_{j,k}$ for all j, k, thus having $\overline{\mathbf{X}}^{[k]}$, $\overline{\mathbf{S}}^{[k]} \in \mathbb{R}^{N \times T}$ for all k.

Like jICA, by estimating a mixing matrix $\overline{\mathbf{A}}^{[k]}$ shared across the concatenated datasets within a given $\overline{\mathbf{X}}^{[k]}$, jIVA estimates a demixing matrix that can better reflect the information shared across the concatenated datasets. Concatenation of datasets also increases the sample size, providing more reliable estimates of underlying sources.

By additionally modeling dependence across different $\overline{\mathbf{X}}^{[k]}$, jIVA can provide improved estimates of sources dependent across the different $\overline{\mathbf{X}}^{[k]}$. This IVA based modeling of dependence provides jIVA with the ability to exploit a powerful yet relaxable assumption of dependence across $\overline{\mathbf{X}}^{[k]}$. In cases where all $\overline{\mathbf{X}}^{[k]}$ are otherwise independent, jIVA becomes equivalent to individual jICA decompositions per each $\overline{\mathbf{X}}^{[k]}$. This additional modeling of dependence across $\overline{\mathbf{X}}^{[k]}$ thus

provides an extra level of diversity which can further help to better estimate sources dependent across $\overline{\mathbf{X}}^{[k]}$.

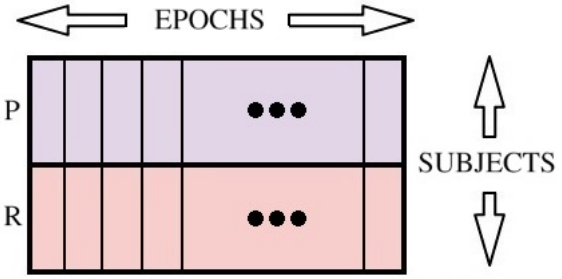


Fig. 1: To construct each data matrix $\overline{\mathbf{X}}^{[k]}$ for a given electrode, measurements across epochs are concatenated horizontally, and measurements across subjects are concatenated vertically. With measurements across subjects over the two conditions ("P" for predictive cue and "R" for random cue), subject measurements across these conditions are also concatenated vertically, so that jIVA may estimate sources that can discriminate between the two conditions.

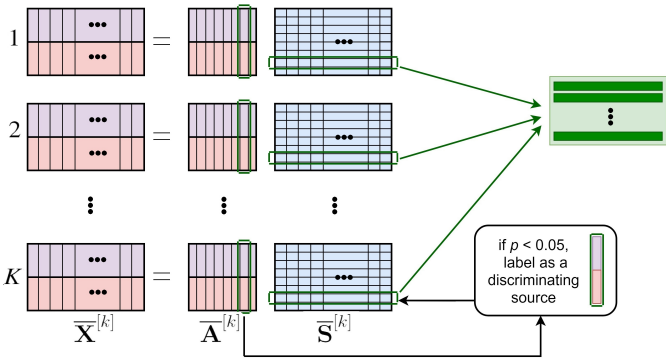


Fig. 2: The jIVA generative model. Applied to EEG, given K electrode datasets of each N mixtures, J epochs are concatenated horizontally to form each data matrix. When using jIVA to discriminate between subject conditions, subject conditions may be concatenated vertically within each data matrix $\overline{\mathbf{X}}^{[k]}$ (e.g. two conditions denoted by purple and red), thus generating sources which are shared or differ between the subject conditions.

By this formulation, jIVA simultaneously leverages dependence across the J concatenated datasets $\mathbf{X}_j^{[k]}$ within a given data matrix $\overline{\mathbf{X}}^{[k]}$, in addition to dependence across the K data matrices, to estimate sources that better reveal information by leveraging use of cross information across all datasets.

In application of jIVA to EEG data, underlying neural processes are measurable across multiple measurement periods (e.g. J epochs), and measurable across multiple electrodes (e.g. K electrodes). Given epoched EEG data described in section II. A., per each electrode, different epoch datasets are reasonably described by some of the same underlying sources, especially those related to the event. By concatenating J epochs into a data matrix $\overline{\mathbf{X}}^{[k]}$ for a single electrode, performing jIVA on these matrices will better reveal information shared across the concatenated epochs within a given $\overline{\mathbf{X}}^{[k]}$, while also providing improved source extraction due to increased sample sizes of source PDFs. Additionally, as sources describing underlying neural processes are naturally dependent across multiple electrodes, e.g. electrodes covering cortical areas related to these processes, jIVA can be used to simultaneously leverage the dependence of measurements across the K relevant electrodes $\overline{\mathbf{X}}^{[k]}$, further improving estimates of sources dependent across electrodes.

While the concatenation of epochs in the jIVA model provides additional statistical power to exploit over the conventional IVA model, concatenation would not be warranted if epoch datasets do not have dependence to exploit. In these situations, decompositions would best be separated on a per-epoch basis as fusion across epochs would be inappropriate. However, as epochs are often repetitions of the same tasks and corresponding measurements, these epochs are well described by the same processes and are thus highly dependent, justifying our model for the application.

E. Discriminating Source Detection after jIVA

Given data recorded across different experimental conditions (e.g. subject condition 1, condition 2, etc.), matrix decompositions such as jIVA can be used to reveal underlying information which aids in discriminating between the conditions. This information may be uncovered by detecting sources which demonstrate statistically significant variability between different conditions in the observed mixtures, and then by interpreting the variability and informative features described by these sources.

Provided with a jIVA decomposition, each row within each source matrix $\overline{\mathbf{S}}^{[k]}$ is a source to be tested for group difference. For the kth data matrix $\overline{\mathbf{X}}^{[k]}$, the ith row of the respective source matrix $\overline{\mathbf{S}}^{[k]}$ is the source which is associated with the ith column of the kth mixing matrix $\overline{\mathbf{A}}^{[k]}$ ($i = 1, \ldots, N$), with this column representing that particular source's contribution to each of the N mixtures of that data matrix $\overline{\mathbf{X}}^{[k]}$. When forming these data matrices, by vertically concatenating each conditions' respective measurements across the same subjects, jIVA may be used to estimate sources that discriminate between the conditions. Discriminatory sources are identifiable when contribution weights (mixing matrix column coefficients) show group difference between conditions. Thus, one may use a statistical test (e.g. paired two-sample t-test) to determine sources with significant difference within a given threshold, e.g., $p = 0.05$. Refer to Fig. 2 to visualize how mixing matrix columns may be used to test for group difference per each respective source row.

In cases when dimension reduction was performed on the data, the demixing matrices estimated by jIVA do not reflect contributions of sources to group-labeled mixtures, and must be "back reconstructed" to reflect these labels. To thus obtain the back-reconstructed mixing matrices in the generative model $\overline{\mathbf{A}}^{[k]}$, we obtain this matrix by multiplying a jIVA estimated mixing matrix $\mathbf{A}^{[k]}$ by the pseudoinverse of the

dimension reduction projection matrix, $\mathbf{R}^{[k]}$, thus giving the generative model mixing matrix $\overline{\mathbf{A}}^{[k]} = \text{pinv}(\mathbf{R}^{[k]})\mathbf{A}^{[k]}$.

III. Simulation Study

Testing the performance of jIVA, we simulated data applicable to the jIVA model, and then compared the inter-symbol interference (ISI) performance of the jIVA estimated solution to that of other data-fusion methods similar to jIVA. ISI is a standard metric for measuring the degree of BSS algorithm performance, measuring the net distortion of a system after estimating the demixing operation $\mathbf{W}$, implying that the mixing operation $\mathbf{A}$ is known. The measured ISI is given by:

$$ISI(\mathbf{G}) = \frac{1}{2N(N-1)}\left[\sum_{n=1}^{N}\left(\sum_{m=1}^{N}\frac{|g_{n,m}|}{\max_p|g_{n,p}|}-1\right) + \sum_{m=1}^{N}\left(\sum_{n=1}^{N}\frac{|g_{n,m}|}{\max_p|g_{p,m}|}-1\right)\right] \quad (6)$$

Here $\mathbf{G} = \mathbf{WA}$ with elements $g_{m,n}$, which in the perfect demixing case is an arbitrarily permuted and scaled identity matrix, corresponding to an ISI value of 0, thus smaller ISI values indicate superior demixing performance. For multiple datasets, ISI can be averaged across the K dataset demixing matrices $\mathbf{W}^{[k]}$ to indicate the separation performance of the entire decomposition [24]. This metric is particularly useful for grading ICA demixing performance when the true mixing is known, such as for simulated studies.

In considering alternative BSS methods comparable to jIVA, such alternatives include performing individual ICAs on each concatenated data matrix $\overline{\mathbf{X}}^{[k]}$, which we refer to here as "joint-ICA" (jICA), or to first average the concatenated epochs within the data matrices $\overline{\mathbf{X}}^{[k]}$ and then perform IVA on these averages (conventional IVA).

The data was generated with 6 datasets and 10 sources per dataset (10 SCVs, 6 sources assigned to each SCV). Across the 10 SCVs, 9 of the SCVs were multivariate Gaussian distributed sources with a random covariance structure, so that these sources featured some level of dependence across the datasets. Within the 10th SCV, we generated 3 of these sources (3 datasets) to be of the same type of sources generated from the other 9 SCVs, while the last 3 datasets' sources of the 10th SCV were generated to resemble EEG signals with an event related signal perturbation. These particular sources were generated as a concatenation of multiple epochs, where each epoch's perturbation event occurred either at the beginning of the epoch or the end of the epoch, so that these sources would be described by 2 clusters of epochs. These perturbation events were aligned for these EEG sources, thus making the sources highly dependent across these last 3 datasets. We chose this last SCV to have 3 of these highly dependent EEG sources to better represent real world situations where dependent sources are not always present across datasets. Source were then mixed by a random mixing matrix per each dataset, with mixing matrix coefficients from the standard normal distribution.

Constructing each epoch to consist of 80 samples, we varied the number of epochs in the data from 10 epochs to 200 epochs (thus also varying the number of samples), and recorded the ISI performance of each method averaged across 100 different simulations with a given number of epochs. Our simulated data is thus the tensor $\mathbf{X}^{[k]} \in \mathbb{R}^{10\times(80*T)}$, $k = 1, \ldots, 6$, where T is equal to 1 for conventional IVA on epoch averages, and equal to the number of total epochs for jICA and jIVA. Fig. 3 plots the mean and standard deviation of ISI values for each method at each number of epochs.

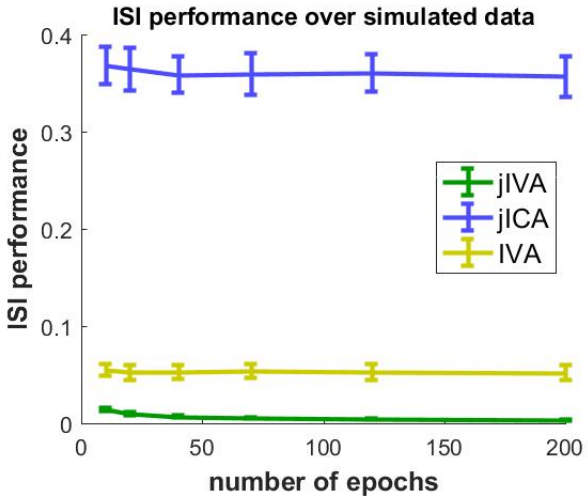


Fig. 3: average ISI performance of three data-fusion methods: jIVA, ICA on each data-matrix ("joint-ICA"), and IVA on averages on epochs ("IVA"). The simulation results show jIVA always outperforming the other two methods, with estimation performance also being more consistent for jIVA.

The ISI performance of jICA is poor relative to the other two methods, as jICA is unable to exploit the source dependence across datasets, which here is particularly needed to separate the Gaussian sources. While IVA and jIVA are able to exploit this dependence, the performance of IVA on the epoch averages is inferior to jIVA because averaging epochs reduces the number of samples within each dataset to the number of samples within an epoch. As the data used 80 samples per epoch, IVA was restricted to estimating source pdfs from only 80 samples, whereas jICA and jIVA estimated source pdfs from 80*(number of epochs) samples per source.

Additionally, the estimated sources of jIVA were able to preserve the per-epoch information better than the other two methods. Instead of averaging the estimated sources' epochs to summarize the epochs, we clustered the epochs and reported cluster centroids, as clustering better reflects the locational variability of the perturbation event, and also produces centroids of estimated sources that can be compared to centroids of the true sources. Thus we divided each source into epochs and clustered a source's epochs using k-means clustering, using silhouette criterion for determining the optimal number of clusters. We reported the cluster centroids as features summarizing the sources, and then compared these centroids of the estimated sources with the centroids of the original sources. Clustering on the jIVA EEG sources resulted in 2

clusters with centroids in agreement with centroids of the original signal's clustering. In comparison, the estimation of the EEG sources were slightly worse for jICA, reflected by poorer quality centroids of the clustered EEG sources. While the performance of conventional IVA was much better than jICA, conventional IVA does not preserve the per-epoch structure of the sources, and thus the source estimated was an average feature that combined features of both clusters, thus yielding a decomposition unable to preserve information on a per-epoch basis.

Having demonstrated the efficacy of jIVA with our simulated data, in the next section we show application to real world EEG data.

IV. Data

A. ARL-BCIT Data

We applied the jIVA model to EEG data obtained from a BCIT study (Brain Computer Interface Technologies) conducted by the Army Research Laboratory in Aberdeen, MD [25], [26]. In the study, 16 male drivers were each fitted with a 64 electrode Biosemi EEG cap, with measurements obtained over a 45-minute simulated driving experiment. The experiment consisted of instances where an audio tone would occur before a perturbation in driving conditions (e.g. simulating wind gusts changing the direction of the car), and drivers were required to adjust their steering in order to correct for changes caused by these perturbations. These instances of tones prior to perturbations were further distinguished between two conditions of tones: tones occurring 1–2 seconds prior to perturbations ("predictive cues"), vs. tones occurring within a random time before, during, or after the perturbations ("random cues"). Using these measurements, we sought to determine whether the jIVA model could generate interpretable features that discriminate measurements over the two conditions.

B. Pre-processing

Recognizing the emerging interest of oscillatory patterns in EEG data (such as theta-band waves and alpha-band waves), for the purpose of characterizing these patterns we pre-process the EEG time-series data to obtain data of EEG band power over time. These oscillatory patterns are researched in previous studies, providing an established benchmark to compare our jIVA results against [25] [26] [27]. Thus prior to jIVA, we first pre-process the time-series data into a band power representation revealing the magnitude as well as the temporal behavior of these patterns.

To pre-process the initial time series data, measurements were first re-referenced to mastoid electrodes, and band pass filtered in the window 0.5 Hz to 50 Hz, for referencing and de-noising of signals. Epochs over each perturbation were then extracted from measurements occurring between 2 seconds before and 4 seconds after the perturbation. Epochs were then down sampled by a factor of 8 (1024 Hz to 128 Hz) and normalized. With each epoch having 768 time series samples, we then generated 640 short-time Fourier transform (STFT) spectrograms of window length 128, spanning all samples of the epoch (i.e. spectrogram n spans samples n to $n+128$, for $1 \leq n \leq 640$). This STFT implementation provides a balance between yielding a high resolution of frequency information and retaining a high number of samples. Within each of these spectrograms, band power measurements were then obtained by summing the bins respective of each band's designated frequency range. Thus, for each epoch, we obtained 640 time series samples of band power measurement for theta band (4-7 Hz), alpha band (8-12 Hz), beta band (13-30 Hz), and gamma band (30-50 Hz). These band power epochs were then separated into predictively cued or randomly cued epochs, based on what type of cue occurred prior to a given epoch.

We then used these band power epochs to form our data matrices for jIVA. With measurements from a given band, measurements across all perturbation epochs (135 epochs) were concatenated together into a single observation, for a given electrode, subject, and cue type. This was under the reasonable assumption that individual epochs are expected to share a common PDF between all other epochs of the same band, electrode, subject, and cue type, and that these epochs were similarly mixed and are thus describable by a common mixing profile. Under these assumptions, concatenation of epochs is justified to leverage the statistical power of jIVA. For jIVA to then be able to estimate sources that discriminate between the different cues, each electrode's data matrix was constructed such that across all subjects, measurements of predictive cues and measurements of random cues were vertically concatenated together, such that the top half and bottom half of a matrix were respectively the predictive cue observations and random cue observations. Thus for a given electrode, a single jIVA data matrix is formed as $\overline{\mathbf{X}}^{[k]} \in \mathbb{R}^{32\times 86400}$, where each of the 32 observations represents either predictive cue or random cue epochs measured from 1 of 16 subjects, and 86400 samples represent 640 samples per epoch across 135 epochs. Fig. 2 explains this construction of an electrode's data matrix from the band power epochs.

C. Dimension reduction and order selection

Real world data often includes measurement noise that may compromise effective source separation, and in general we have an overdetermined problem where there are more observations than informative signals (sources). Thus, for the electrode data matrices $\overline{\mathbf{X}}^{[k]}$, we performed dimension reduction using principal component analysis (PCA) prior to jIVA. As jIVA requires each data matrix $\overline{\mathbf{X}}^{[k]}$ to have the same number of mixtures, we sought an effective common order $\widetilde{P}$ to use across the jIVA grouped electrodes.

We used several techniques for estimating the source order $\widetilde{P}$, including AIC, MDL, and methods to take dependent samples into account in the likelihood formulation [28] [29] [30], these techniques varied considerably in estimated order accuracy for a given data matrix $\overline{\mathbf{X}}^{[k]}$, and varied considerably in orders chosen across different $\overline{\mathbf{X}}^{[k]}$. When such order estimation techniques fail to converge upon a commonly agreed upon order, it is justified to examine decompositions across a range of order candidates, and then use an order

that gives the best performance (in terms of e.g. stability of results, and ability to identify electrodes with sources that discriminate between cues). Thus, we ultimately used the order $\widetilde{P}$ that revealed the highest number of electrodes with a source discriminating between the cues. This entailed performing PCA dimension reduction on each electrode data matrix $\overline{\mathbf{X}}^{[k]}$, using a given order candidate, then performing individual jICA decompositions on each dimension reduced data matrix, and then determining if a given electrode's sources are discriminatory using the paired two sample t-test explained in section II. part E. For all bands analyzed, the highest number of electrodes with discriminating sources was found with an order of 10, and thus we chose this order for dimension reduction of each $\overline{\mathbf{X}}^{[k]}$ prior to jIVA. We gained further confidence in this order by determining that jICA and jIVA sources were consistently estimated across multiple runs, thus ensuring the stability of results.

D. Electrode groupings for jIVA

In order for jIVA to effectively leverage cross information across data matrices $\overline{\mathbf{X}}^{[k]}$, we then sought sensible groupings of electrodes that share discriminatory information. Using our chosen dimension reduction order, we determined which jlCA-sourced electrodes individually discriminated between conditions. Then, using a map of individual-discriminating electrodes, a neuroscientist selected anatomic areas of interest related to the task. Using groups of electrodes within those anatomic areas, we tested jIVA for its ability to generate discriminatory sources. Fig. 4 shows the map of individual-discriminating electrodes from jICA, and the selected electrode groupings for jIVA.

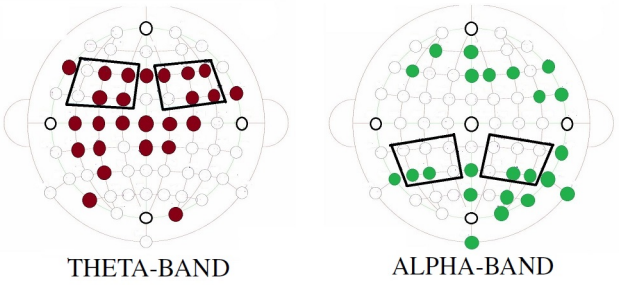


Fig. 4: Solid circles denote electrodes with 1 or more jICA sources showing group difference ($p < 0.05$), using PCA dimension reduction order 10. With this mapping, 4 regions for jIVA (indicated above) were chosen from the electrodes, grouping 6 electrodes per each region ("theta-left", "theta-right", "alpha-left", "alpha-right").

As interest has been given to alpha activity in the visual cortex resulting from task related events, we chose two regions of six electrodes each for alpha band, symmetric across the cortex ("alpha-left", "alpha-right"). Similarly for theta band, as the majority of discriminatory source activity appeared to be prefrontal, we chose two regions of six electrodes each in the prefrontal region ("theta-left", "theta-right"). While beta band and gamma band also demonstrated jICA-sourced electrodes that discriminated between the conditions, the discriminating electrodes for these bands were much fewer in number and were not localized to specific regions. As we primarily expect to see differences in states within theta and alpha band, we constructed jIVA groupings only for the theta and alpha bands. We supported our decision to only use theta and alpha bands from further analysis showing that discriminatory activity in beta and gamma bands was not present.

Our full electrode-grouping structure for jIVA is thus given by $\overline{\mathbf{X}}^{[k]} \in \mathbb{R}^{32\times 86400}$, for $k = 1, \ldots, 6$, for each selected region of 6 electrodes.

E. jIVA using IVA-GGD

Having electrode groupings ready for our jIVA model, we use IVA-GGD as our implementation of jIVA. IVA-GGD assumes a multivariate generalized Gaussian PDF model [31], which is appropriate for this data, as this is a sufficiently general model covering the sub-Gaussian and super-Gaussian source marginals. We thus applied jIVA to the data using IVA-GGD, and then determined the jIVA sources discriminating between conditions by using a paired two sample t-test ($p < 0.05$) on each source's respective mixing matrix column. To acquire informative features from these discriminating sources (with each source being 135 epochs of each 640 samples), we then processed information across epochs, via grouping of epochs using clustering.

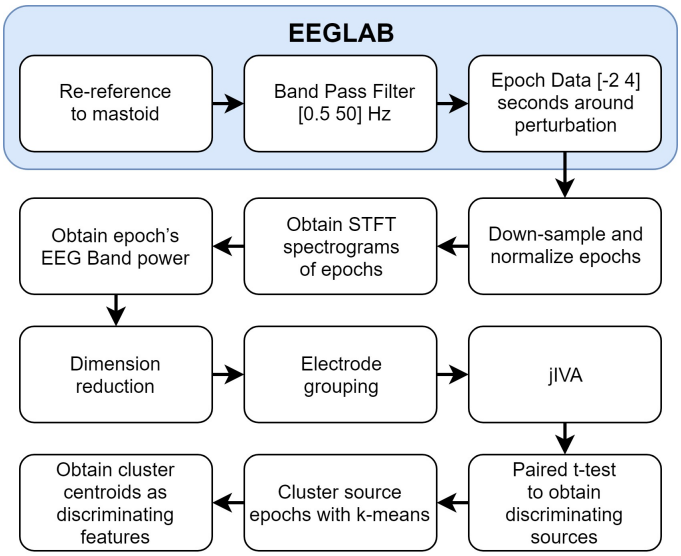


Fig. 5: Summary of steps used to obtain discriminating features with jIVA, given the original EEG data.

V. Summarizing Sources via Clustering

Within each significant discriminating source estimated from jIVA, epochs exhibited considerable variability. Since averaging would cause loss of information given the variability in source estimates corresponding to each epoch, clustering the epochs and reporting cluster centroids as informative features

instead provides an effective means to summarize the epochs while also adequately reflecting their variability.

To cluster the epochs of a single source, we used k-means clustering, with Pearson correlation as the distance metric. We used silhouette criterion to determine the optimal number of clusters [32], after determining that between various clustering evaluation criteria (silhouette, Davies-Bouldin, Calinski-Harabasz. and gap), all criteria on average agreed the most with silhouette. In addition to using k-means, we also compared k-means results to those of hierarchical clustering and affinity propagation clustering. For all clustering algorithms, we used silhouette criterion across multiple clustering solutions to determine the optimal number of clusters, and used correlation as the distance metric. Results from the other two clustering methods consistently compared very similarly to results of k-means (via number of clusters, cluster sizes, and centroid epoch shapes), thus all clustering methods closely agreed with each other. For the following results, we used k-means to cluster the discriminating sources, and report the centroids of these sources as features useful for discrimination between conditions.

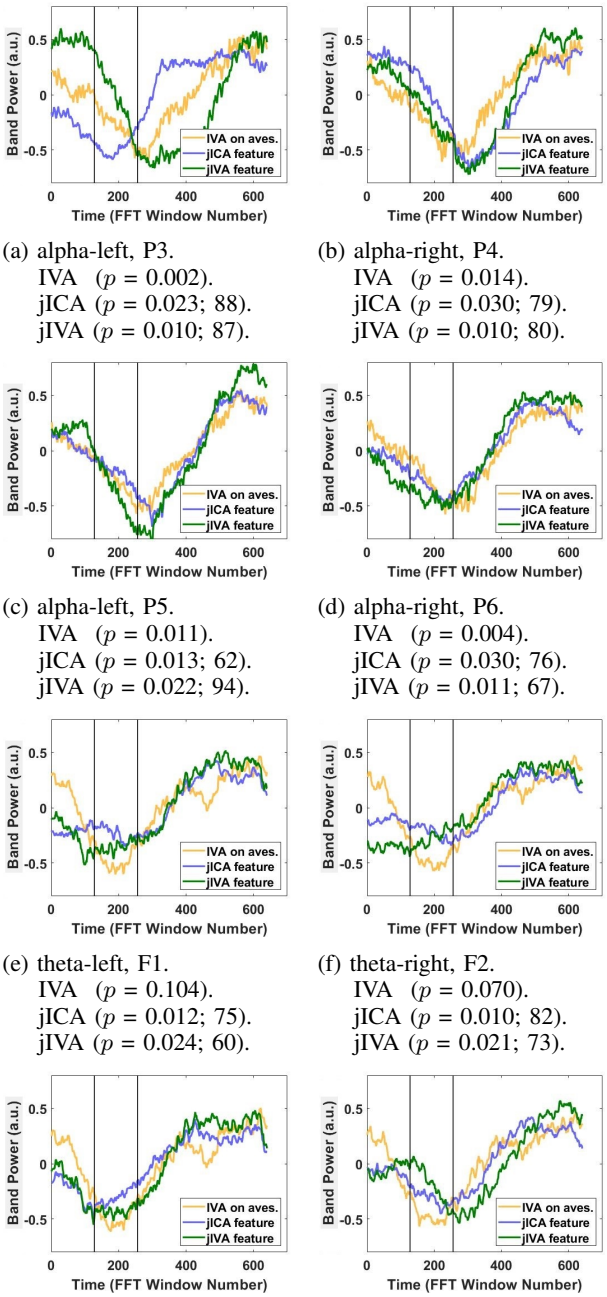
VI. Results and Discussion

Across the four electrode groupings we applied the jIVA method on ("theta-left", "theta-right", "alpha-left", "alpha-right"), all groupings produce discriminating sources across multiple electrodes, all aligned within the same SCV. Thus, by detecting the discriminating information dependent across the electrodes, jIVA is able to exploit this dependency to better estimate sources describing the discriminating information.

To evaluate the efficacy of jIVA for obtaining discriminating features, we compare jIVA features to features obtained from two similar BSS methods: jICA applied on individual electrode data matrices ("jICA", $\overline{\mathbf{X}} \in \mathbb{R}^{32\times 86400}$), and IVA performed on the averages of concatenated epochs ("IVA on epoch averages," or simply "IVA", $\mathbf{X}^{[k]} \in \mathbb{R}^{32\times 640}$, $k = 1, \ldots, 6$). As IVA models dependence across electrodes without modeling dependence across epochs, and as jICA models dependence across epochs without modeling dependency across electrodes, comparing jIVA to these two methods demonstrates the additional statistical power of jIVA by modeling both dependencies.

For ease of jointly interpreting the discriminating sources produced across the different BSS methods ("IVA," "jICA," "jIVA"), we orient all sources such that the displayed features represent a greater activation for predictive cues. This entails flipping the orientation of sources (multiplication by -1), if the source shows greater average activation for random cues (i.e., if the mean of mixing matrix coefficients for random cues is greater than the respective mean for predictive). We then cluster these sources and report centroids as the discriminating features (for jICA and jIVA), or report the source itself as the discriminating feature (for IVA).

The common pattern consistent across all methods is an event related decrease in band power, characterized by a decrease in theta/alpha power near the perturbation, followed



(a) alpha-left, P3.
IVA (p = 0.002).
jICA (p = 0.023; 88).
jIVA (p = 0.010; 87).

(b) alpha-right, P4.
IVA (p = 0.014).
jICA (p = 0.030; 79).
jIVA (p = 0.010; 80).

(c) alpha-left, P5.
IVA (p = 0.011).
jICA (p = 0.013; 62).
jIVA (p = 0.022; 94).

(d) alpha-right, P6.
IVA (p = 0.004).
jICA (p = 0.030; 76).
jIVA (p = 0.011; 67).

(e) theta-left, F1.
IVA (p = 0.104).
jICA (p = 0.012; 75).
jIVA (p = 0.024; 60).

(f) theta-right, F2.
IVA (p = 0.070).
jICA (p = 0.010; 82).
jIVA (p = 0.021; 73).

(g) theta-left, F3.
IVA (p = 0.140).
jICA (p = 0.010; 53).
jIVA (p = 0.015; 55).

(h) theta-right, F4.
IVA (p = 0.065).
jICA (p = 0.005; 71).
jIVA (p = 0.011; 68).

Fig. 6: Comparison of features from three BSS methods ("IVA on averages, i.e. "IVA", "jICA", "jIVA"). With the perturbation at time-series sample 256, and as the FFT window length is 128 samples, the perturbation is thus observed between FFT window 128 and 256, denoted by vertical bars. p-values for group difference significance are given for each feature, and cluster centroid features (jICA and jIVA) have reported the number of epochs assigned to that centroid's cluster, out of 135 total epochs. (e.g. p = 0.022; 94 epochs). The p-values are uncorrected for multiple comparisons.

by an increase in theta/alpha power immediately after perturbation. When interpreting results of epoch clustering, we put the emphasis on cluster centroids that show the characteristic theta/alpha decrease near perturbation, regardless of the size of clusters (all clusters with centroids having this shape contain at least 50 epochs assigned to that cluster). For all discriminating sources post-processed by clustering, silhouette criteria for clustering evaluation determines that the optimal number of clusters is most often 2 clusters.

A. Comparing jIVA to jICA and IVA

Fig. 6 compares the jIVA features to respective features from jICA and IVA, for various electrodes across our jIVA designated electrode groupings.

IVA on epoch averages (labeled as "IVA on aves.") entails performing IVA on the jIVA electrode groupings, but instead of horizontally concatenating the epochs, the epochs are averaged together for each concatenated row and IVA is performed on these epoch averages ($\mathbf{X}^{[k]} \in \mathbb{R}^{32\times 640}, k = 1, \ldots, 6$). IVA on epoch averages also delivers discriminating sources with statistical significance, and like jIVA, these discriminating sources across electrodes are all assigned to the same SCV. As this method averages the concatenated epochs together prior to IVA, the method has fewer samples to use for estimating source PDFs and is thus expected to suffer over methods leveraging larger numbers of samples. In addition, because the estimated sources of this method are averages over epochs, whereas jICA and jIVA sources are concatenations of epochs, jICA and jIVA sources provide the advantage of preserving information on a per epoch basis. While IVA on epoch averages provides discriminating features for alpha band that are interpretable, the discriminating features that IVA on epoch averages estimates for theta are hard to interpret due to their noisy nature. Additionally, while IVA on epoch averages does estimate interpretable theta band components with the characteristic shape of band power decrease at perturbation, none of these components discriminate under the threshold $p < 0.05$ (as shown in Fig. 6).

Joint-ICA applied on individual electrode data matrices ($\overline{\mathbf{X}} \in \mathbb{R}^{32\times 86400}$) entails individual jICA decompositions per each electrode's data matrix from a jIVA electrode grouping. Compared to IVA on epoch averages, jICA is able to obtain interpretable discriminating features for theta band, which IVA on epoch averages is unable to do. However, compared to our jIVA method, jICA tends to yield fewer electrodes with disciminating sources (e.g. jICA of "alpha-left" yields 2 group discriminating-sourced electrodes, while jIVA of "alpha-left" yields 4; jICA of "alpha-right" yields 2, while jIVA of "alpha-right" yields 6). This demonstrates that jICA suffers by not using cross-information between electrodes to aid in extracting discriminating information, as jIVA is instead able to identify more discriminating sources from additional electrodes.

Joint-IVA demonstrates the previously noted advantages seen in both IVA and jICA. Like jICA, jIVA can obtain interpretable discriminating features for the theta band, and can provide sources that preserve variability on a per epoch basis (as opposed to IVA sources). Like IVA, jIVA can leverage dependence across electrodes to better estimate discriminating processes dependent across electrodes, providing a decomposition where the discriminatory processes can be linked across electrodes (via source alignment), and providing a decomposition identifying additional electrodes that show discrimination between the cues (as opposed to jICA). While the jICA type data definition provides a good model match across epoch measurements of a given electrode, this type of modeling imposes a strong assumption on the data, and thus is a more constrained model on the data. By the additional modeling of dependence across electrodes via IVA, jIVA allows a relaxation on the strong assumptions imposed through jICA, thus providing an appropriate balance between flexibility and robustness on the resulting decomposition.

B. Interpreting the jIVA features

To understand the interpretive value of the jIVA features, it is first necessary to understand the structure of the data within the jIVA model. For the given bands analyzed, theta and alpha, both of these bands are markers of a number of processes including executive control, action planning, sensory gating, visual attention allocation, spatial processing, among several others depending on location of the activity and of the task [33], [34]. In this task, as a perturbation occurs within an observed epoch, subject disinhibition in reaction to the perturbation generally entails a localized drop in theta/alpha band power as the subject responds to the perturbation, and a subsequent increase in theta/alpha band power once the subject response completes. Within a given jIVA data matrix $\overline{\mathbf{X}}^{[k]}$, any observed epoch demonstrates some variation of this explained decrease/increase in band power. Thus, jIVA sources can be understood as providing a basis encompassing patterns of band power change over the observed epoch measurements, providing a means to explain the structural variation of the observed epochs. In this work, significant differences in neural activity based on predictive versus random cuing were observed both in the frontal theta and parietal alpha frequency bands.

The jIVA discriminating features may be interpreted as summarizing a form of band power change pattern that differs between the two cue types. An understanding of the difference between cues may help to explain the shape/structure of these features. As these discriminating features have higher activation in the predictive cued, the shape of the discriminating features may be tied to how the predictively cued have prior knowledge of the perturbation, and may reflect the brain efficiently operating on that prior knowledge (e.g. an earlier timing of local minimum may reflect an earlier and perhaps more efficient correction in response to perturbation).

An improved understanding of these features may be developed by additionally comparing the discriminating features with those that do not discriminate, within a given jIVA data matrix decomposition. In our analysis, sources that did not discriminate (usually 9 out of 10 sources per data matrix) had noisy characteristics and generally do not follow a prominent trend in band power change with respect to the perturbation event. In contrast, the discriminating features show a relatively "noise-free" and prominent trend of band power increase prior

to the perturbation, and decrease immediately after. Hence the statistically significant components demonstrate a more consistent and prominent trend of increase or decrease in band power relative to the timing of the perturbation.

Across the discriminating features, some notable observations can be made. Generally across theta and alpha bands, the components that were significantly more prominent in the predictively cued show a decrease in band power within the perturbation window (theta) and just after the perturbation window (alpha). Theta activity has been linked to well-established role in executive control functions [35], [36] and is frequently related to response inhibition [37]. In this context, it appears as the jIVA theta features potentially reflect differences in control strategies in the predictive vs random cue conditions.

With respect to the alpha related activity, there is a clear desynchronization that occurs peri-perturbation which may reflect modulation in sensory and/or motor processing. One rhythm that shows similar activity is the well-studied "mu" rhythm that has been related to sensorimotor transformations wherein visual (and other sensory information) are translated into action [38], [39]. It is difficult in this context to ascribe these changes in ongoing alpha oscillations to specific mu rhythms since the latter is specific to motor cortex and movement, but it seems plausible in this task that the desynchronization observed peri-perturbation is related to impending visuomotor processing required to perform visuomotor processing in order to respond to the perturbation. In this particular experiment, this processing appears to be more prominent in the predictive cueing conditions.

VII. Conclusion

In this paper, we introduce an IVA model with a jICA-type data definition, which we term jIVA, and we demonstrate how this model can be applied to obtain discriminating features from EEG datasets.

In comparing jIVA to IVA implemented on the averages of concatenated epochs, we demonstrate that jIVA can leverage increased statistical power, via large sample properties and preservation of epoch variability, to provide more informative features describing group difference between conditions. We also show that jIVA is able to provide meaningful discriminating features in cases when IVA on concatenated epoch averages cannot, as seen for theta band electrode groupings.

In comparing jIVA to jICA implemented on individual electrode concatenated datasets, we demonstrate that jIVA can leverage dependence across electrodes to better estimate discriminatory information present across electrodes. Unlike jICA, jIVA provides the additional benefit of linking processes dependent across electrodes via source alignment. Across all four jIVA electrode groupings, jIVA additionally identifies more electrodes than jICA that discriminate between cues.

Thus, the modeling of dependence across concatenated datasets and across data matrices provides jIVA with additional statistical power for more robust source separation.

References

[1] A. Hyvärinen and E. Oja, "Independent component analysis: algorithms and applications," *Neural networks*, vol. 13, no. 4-5, pp. 411–430, 2000.

[2] S. Makeig *et al.*, "Independent component analysis of electroencephalographic data," in *Advances in neural information processing systems*, 1996, pp. 145–151.

[3] G. R. Naik and D. K. Kumar, "An overview of independent component analysis and its applications," *Informatica*, 2011.

[4] A. X. Stewart *et al.*, "Single-trial classification of EEG in a visual object task using ICA and machine learning," *Journal of neuroscience methods*, vol. 228, pp. 1–14, 2014.

[5] P. P. Acharjee *et al.*, "Independent vector analysis for gradient artifact removal in concurrent EEG-fMRI data."

[6] D. K. Emge *et al.*, "Independent vector analysis for SSVEP signal enhancement," in *Information Sciences and Systems (CISS), 2015 49th Annual Conference on*. IEEE, 2015, pp. 1–6.

[7] A. Delorme *et al.*, "Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis," *Neuroimage*, vol. 34, no. 4, pp. 1443–1449, 2007.

[8] J. Laney *et al.*, "Capturing subject variability in fMRI data: a graph-theoretical analysis of GICA vs. IVA," *Journal of neuroscience methods*, vol. 247, pp. 32–40, 2015.

[9] J. Sui *et al.*, "A review of multivariate methods for multimodal fusion of brain imaging data," *Journal of neuroscience methods*, vol. 204, no. 1, pp. 68–81, 2012.

[10] T. Adali *et al.*, "Multimodal data fusion using source separation: Two effective models based on ICA and IVA and their properties," *Proceedings of the IEEE*, vol. 103, no. 9, pp. 1478–1493, 2015.

[11] ——, "Multimodal data fusion using source separation: Application to medical imaging," *Proceedings of the IEEE*, vol. 103, no. 9, pp. 1494–1506, 2015.

[12] M. A. Akhonda *et al.*, "Consecutive independence and correlation transform for multimodal fusion: Application to EEG and fMRI data," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2018, pp. 2311–2315.

[13] S. Bhinge *et al.*, "Non-orthogonal constrained independent vector analysis: Application to data fusion," in *Acoustics, Speech and Signal Processing (ICASSP), 2017 IEEE International Conference on*. IEEE, 2017, pp. 2666–2670.

[14] X. Chen *et al.*, "Joint blind source separation for neurophysiological data analysis: Multiset and multimodal methods," *IEEE Signal Processing Magazine*, vol. 33, no. 3, pp. 86–107, 2016.

[15] V. D. Calhoun *et al.*, "Neuronal chronometry of target detection: fusion of hemodynamic and event-related potential data," *Neuroimage*, vol. 30, no. 2, pp. 544–553, 2006.

[16] T. Kim *et al.*, "Independent vector analysis: An extension of ICA to multivariate components," in *International Conference on Independent Component Analysis and Signal Separation*. Springer, 2006, pp. 165–172.

[17] J. Sui *et al.*, "An ICA-based method for the identification of optimal fMRI features and components using combined group-discriminative techniques," *Neuroimage*, vol. 46, no. 1, pp. 73–86, 2009.

[18] V. D. Calhoun *et al.*, "Method for multimodal analysis of independent source differences in schizophrenia: combining gray matter structural and auditory oddball functional data," *Human brain mapping*, vol. 27, no. 1, pp. 47–62, 2006.

[19] J. Laney *et al.*, "Quantifying motor recovery after stroke using independent vector analysis and graph-theoretical analysis," *NeuroImage: Clinical*, vol. 8, pp. 298–304, 2015.

[20] J. Brooks *et al.*, "Differential functionality of right and left parietal activity in controlling a motor vehicle," *Frontiers in systems neuroscience*, vol. 10, p. 106, 2016.

[21] P. Ghorbanian *et al.*, "Identification of resting and active state EEG features of alzheimer's disease using discrete wavelet transform," *Annals of biomedical engineering*, vol. 41, no. 6, pp. 1243–1257, 2013.

[22] T. Adali *et al.*, "Diversity in independent component and vector analyses: Identifiability, algorithms, and applications in medical imaging," *IEEE Signal Processing Magazine*, vol. 31, no. 3, pp. 18–33, 2014.

[23] ——, "IVA and ICA: Use of diversity in independent decompositions," in *Signal Processing Conference (EUSIPCO), 2012 Proceedings of the 20th European*. IEEE, 2012, pp. 61–65.

[24] O. Macchi and E. Moreau, "Self-adaptive source separation by direct and recursive networks," in *Proc. International Conference on Digital Signal Processing (DSP'93) Limasol, Cyprus*, 1993.

[25] J. Brooks and S. Kerick, "Event-related alpha perturbations related to the scaling of steering wheel corrections," *Physiology & behavior*, vol. 149, pp. 287–293, 2015.
[26] J. O. Garcia *et al.*, "Estimating direction in brain-behavior interactions: Proactive and reactive brain states in driving," *NeuroImage*, vol. 150, pp. 239–249, 2017.
[27] J. A. Meltzer *et al.*, "Individual differences in EEG theta and alpha dynamics during working memory correlate with fMRI responses across subjects," *Clinical Neurophysiology*, vol. 118, no. 11, pp. 2419–2436, 2007.
[28] G.-S. Fu *et al.*, "Likelihood estimators for dependent samples and their application to order detection," *IEEE Transactions on Signal Processing*, vol. 62, no. 16, pp. 4237–4244, 2014.
[29] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 33, no. 2, pp. 387–392, 1985.
[30] Y.-O. Li *et al.*, "Estimating the number of independent components for functional magnetic resonance imaging data," *Human brain mapping*, vol. 28, no. 11, pp. 1251–1266, 2007.
[31] M. Anderson *et al.*, "Independent vector analysis: Identification conditions and performance bounds," *IEEE Transactions on Signal Processing*, vol. 62, no. 17, pp. 4399–4410, 2014.
[32] P. J. Rousseeuw, "Silhouettes: a graphical aid to the interpretation and validation of cluster analysis," *Journal of computational and applied mathematics*, vol. 20, pp. 53–65, 1987.
[33] W. Klimesch *et al.*, "EEG alpha oscillations: the inhibition–timing hypothesis," *Brain research reviews*, vol. 53, no. 1, pp. 63–88, 2007.
[34] M. Kawamata *et al.*, "Event-related desynchronization of frontal-midline theta rhythm during preconscious auditory oddball processing," *Clinical EEG and neuroscience*, vol. 38, no. 4, pp. 193–202, 2007.
[35] Enriquez-Geppert *et al.*, "Modulation of frontal-midline theta by neurofeedback," *Biological psychology*, vol. 95, pp. 59–69, 2014.
[36] ——, "EEG-neurofeedback as a tool to modulate cognition and behavior: a review tutorial," *Frontiers in human neuroscience*, vol. 11, p. 51, 2017.
[37] J. F. Cavanagh and M. J. Frank, "Frontal theta as a mechanism for cognitive control," *Trends in cognitive sciences*, vol. 18, no. 8, pp. 414–421, 2014.
[38] J. A. Pineda, "The functional significance of mu rhythms: translating "seeing" and "hearing" into "doing"," *Brain research reviews*, vol. 50, no. 1, pp. 57–68, 2005.
[39] M. Sabate *et al.*, "Mu rhythm, visual processing and motor control," *Clinical Neurophysiology*, vol. 123, no. 3, pp. 550–557, 2012.