ICA in Electroencephalography: A Small Pipeline Study

John Doe

Supervisor:

2020, June

Abstract

This paper has two aims: a) Examine whether or not changes in select hyper-parameters can influence the perceived performance of different ICA methods in our pipelines. b) Assess the performance of different ICA methods in different pipelines on Electroencephalography (EEG) data from the BCI competition IV [Brunnder et al., 2008].

In the paper we show that fastICA leads to the best performance while pipelines with SOBI consistently performs the worst out of our considered ICA methods. Further we will establish that some choices in hyper-parameters influence the performance of the pipeline very little, e.g. band-pass filtering data in the β-range (13-30 Hz) contra band-pass filtering data in the α- (8-13 Hz) and β-range. Other options such as classification algorithm seem to have a greater impact on performance

Contents

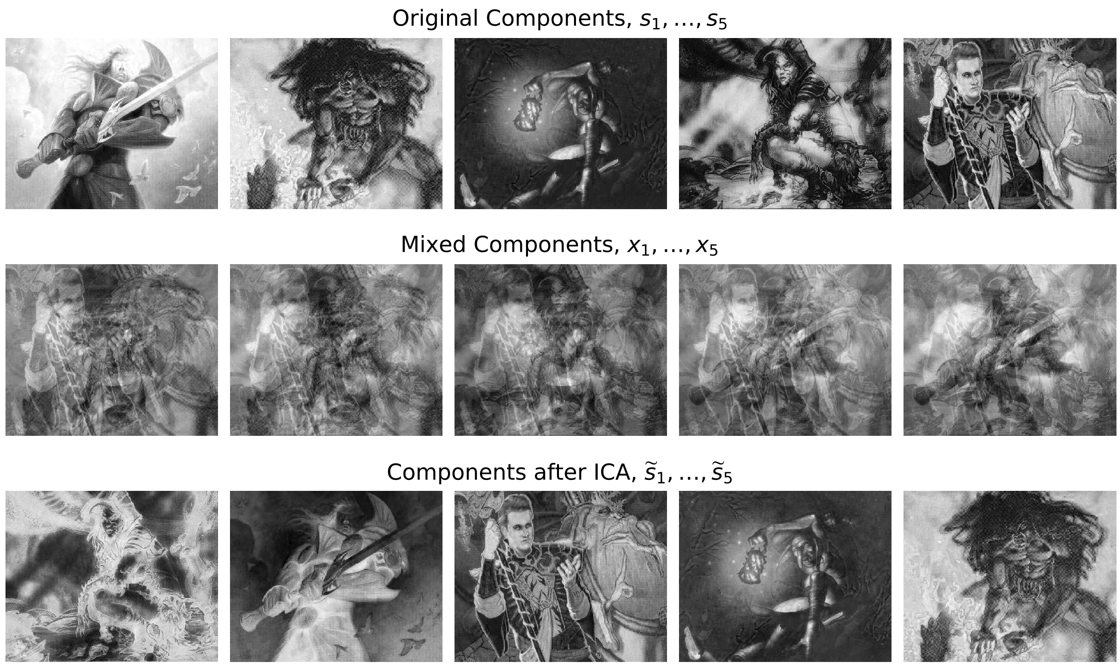

Figure 1: The pictures are not in order anymore and two of the pictures are the negative of the original. This is be elaborated on in Section 2.3. All rights reserved Wizards of the Coast.

1 Introduction

Consider the following scenario: You are an avid Magic The Gathering (MTG) player. One of the things you love the most about the game is the art on the cards. You have therefore downloaded the art of 5 of your favorite MTG cards. To your horror you realize one day that all the pictures are mixed as in Figure 1. Unfortunately you cannot remember where you obtained the pictures in the first place so the question thus becomes: *Can the pictures be unmixed*?

The answer turns out to be yes. A picture can be represented by a sequence of numbers, one for each pixel, signifying the color of the given pixel. This way each of the pictures in the second row of Figure 1, denoted by x_i, $i = 1, \ldots, 5$, can be considered as a linear mix of the five original pictures denoted by s_i. With this we have for each pixel t:

$$
\begin{aligned}
x_1(t) &= a_{11}s_1(t) + a_{12}s_2(t) + \ldots + a_{15}s_5(t) \\
x_2(t) &= a_{21}s_1(t) + a_{22}s_2(t) + \ldots + a_{25}s_5(t) \\
&\vdots \\
x_5(t) &= a_{51}s_1(t) + a_{52}s_2(t) + \ldots + a_{55}s_5(t).
\end{aligned}
$$

where a_{ij} $i, j = 1, \ldots, 5$ are some real-valued parameters the computer mixed each picture with. If we knew the a_{ij} parameters our job would be done as our problem would be equivalent to solving five linear equations with five unknowns, but the snag is precisely that the *parameters are unknown*.

It turns that out under some rather mild assumptions it is possible to recover the original sources $\widetilde{s}_i$ with *Independent Component Analysis*, henceforth ICA for short. The result of ICA can be seen in the third row of pictures in Figure 1.

ICA has over the last 40 years established itself as a mature field of research with a wide array of applications, with more direct usefulness than unmixing MTG card art, from econometrics to recordings of the brain [Hyvärinen and Oja, 2000, pp.11-12]. More precisely we will be analyzing electroencephalographic (EEG) data. EEG data consists of recordings of electrical potentials in different locations on the scalp. Supposedly these potentials are generated by a mixture of underlying brain and muscle activity [Hyvärinen and Oja, 2000, p.150]. The problem is similar to the one above: We would like to know the original sources, but the mixing of the sources is unknown.

1.1 Outline of Objective We will in this paper test different ICA methods by comparing the performance achieved in different pipeline setups. In this paper performance is measured by a classifiers ability to accurately guess certain motor imagery actions using features extracted from the individual unmixed signals from the ICA's.

When assessing accuracy in experimental setups like these, there are many choices to be made in regards to parameters and hyper-parameters. It is clear that different choices can lead to different conclusions. The question of scientific objectivity is not new and it is hardly positioned to solved anytime soon[1].

But no matter if you are standing in the Popperian or the Kuhnian corner, it is important to be aware of how otherwise sensible choices can color our perception of different models and their mutual strength. This paper is therefore in part a small scale attempt to do just that. By exploring different parameter settings we will examine whether or not the relative performance of different ICA's is consistent across different pipelines. In the end we will hopefully be able to identify which ICA performs the best on our EEG data. In summary our two overarching goals are:

a) Examine if different choices in hyper-parameters result in different relative score.
b) Assess the performance of different ICA methods on EEG data.

2 Theory

In this section we outline our ICA framework and introduce the different ICA's in consideration. We consider from the start the *noisy ICA* setting[2]. The basic structure follows [Hyvärinen and Oja, 2000].

[1] see e.g. [Chakravartty, 2017] for an overview of the debate.

[2] See e.g. [Hyvärinen et al., 2001] for this nomenclature, although the assumptions on the noise are more restrictive.

2.1 Towards Rigidity Assume we observe n linear combinations of n independent components with added noise variable $\boldsymbol{\epsilon}$

$$x_i = a_{i1}s_1 + a_{i2}s_2 + \ldots + a_{in}s_n + \epsilon_i, \quad \forall\, i \in \{1, 2, ..., n\}.$$

Strictly speaking we should have a time index t in order to differentiate between each time instant, but we have suppressed it for simplicity.

We consider each x_i, s_i, ϵ_i as random variables. More succinctly we can consider the random vectors for each time instant $\mathbf{x} = (x_1, x_2, \ldots x_n)$, $\mathbf{s} = (s_1, s_2, \ldots, s_n)$, the invertible square mixing matrix A with entrances a_{ij} and noise vector $\boldsymbol{\epsilon} = (\epsilon_1, \epsilon_2, \ldots, \epsilon_n)$. We can define the ICA model as

$$\mathbf{x} = \mathbf{A}\mathbf{s} + \boldsymbol{\epsilon}.$$

We do not make any assumption on the noise term $\boldsymbol{\epsilon}$ in our initial setup as all the different algorithms under consideration impose different restrictions on it which will be made more explicit later.

The base assumption in ICA is then that the original components are independent, i.e. $s_i \perp\!\!\!\perp s_j$ for $i \neq j$. From this an unmixing matrix A^{-1} which maximizes independence between the signals it recovers is to be determined in order to compute the noisy sources

$$\mathbf{A}^{-1}\mathbf{x} = \mathbf{A}^{-1}(\mathbf{A}\mathbf{s} + \boldsymbol{\epsilon}) = \mathbf{s} + \mathbf{A}^{-1}\boldsymbol{\epsilon} = \hat{\mathbf{s}}.$$

One should here not be confused between *sensor noise and source noise*, i.e. the difference between

$$\mathbf{x} = \mathbf{A}\mathbf{s} + \boldsymbol{\epsilon} \quad \text{and} \quad \mathbf{x} = \mathbf{A}(\mathbf{s} + \boldsymbol{\epsilon}).$$

The noise term in the second equation can be described as source noise where the noise is added to the independent components as a whole, which is in contrast to the first equation where the noise is added to *each independent component separately*. The first equation is the noisy ICA setting which we have described, whereas the second problem simplifies to the normal ICA setting with $\hat{\mathbf{s}} = s + \boldsymbol{\epsilon}$ (cf. [Hyvärinen and Oja, 2000, p.94]).

In the preceding paragraphs we assumed the number of observed sources was equal to the number of original sources. This assumption can be somewhat relaxed as one can also have an undercomplete ICA's, meaning there are *more* observed sources than independent components.

The reverse problem of overcomplete ICA's where one has a higher number of original sources than observed sources is harder. We will not pursue the problem in this paper, but the reader is referred to [Hyvärinen and Oja, 2000, ch.16].

2.2 Free Appetizers Before embarking on the actual ICA algorithms one often wants to do some preprocessing on the data. The two most important preprocessing steps for our present context are *band-pass filtering* and *whitening*.

Band-pass filtering is done to suppress unwanted artifacts from the data. In our EEG context one may wanted to band-pass filter the data since it is known that brainwaves only exists on certain frequencies. One could therefore, in theory, filter away non-brainwaves signals.

The other important preprocessing step is whitening, sometimes interchangeably referred to as PCA although PCA is in this context just a *type* of whitening. The idea of whitening is to transform ones data such that all components are uncorrelated with each other and their covariance equals the identity. In the non-noisy ICA setting this process reduces the ICA search space in addition to often reducing noise and preventing over-fitting and it is therefore a standard preprocessing step in fastICA [Hyvärinen and Oja, 2000, p.12]. However, whitening has some limitations in a noisy ICA setting.

2.2.1 Whitening in a Noisy World We have illustrated the limitations of whitening in a noisy ICA setting in Figure 2. On the first we row in the first panel we see the original two components s_1 and s_2 which are each sampled from two i.i.d. uniform distributions.

The next panel show the mixed components in the non-noisy ICA setting. The components are no longer independent and they now form a parallelogram which seems to have a direction. In the last panel on the fist row we see the mixed components in the noisy ICA setting using the same mixing matrix. We have sampled the noise from a multivariate normal distribution of the form

$$\epsilon \sim \mathcal{N}\left(\begin{pmatrix} 5 \\ 5 \end{pmatrix}, \begin{pmatrix} 5 & 0 \\ 0 & 5 \end{pmatrix}\right).$$

We still see the dependence in the joint distribution, but the direction is not as conspicuous as in the previous plot.

In the first panel on the second row we have performed whitening in the non-noisy ICA. We see whitening has completely recovered the shape of the original distribution. The only thing left is calculating the angle with which the whitened distribution has been rotated wrt. the original distribution.

Another way to realize this is by considering the dimension of the search space. A n dimensional mixing matrix has n^2 degrees of freedom. After whitening one only needs to estimate an orthogonal unmixing matrix which has $n(n-1)/2$ degrees of freedom or about half as many [Hyvärinen et al., 2001, p.160]. In the 2-dimensional case this amounts to finding a single angle parameter.

In the second panel on the second row we see whitening performed in the noisy ICA setting. We see that the original shape of the distribution has not been recovered and thus that our search space has not been reduced to just estimating an orthogonal unmixing matrix.

The last panel shows the same noisy ICA setting whitened, but now we have corrected for the noise. This way we have obtained the exact same rotated distribution as in the non-noisy ICA

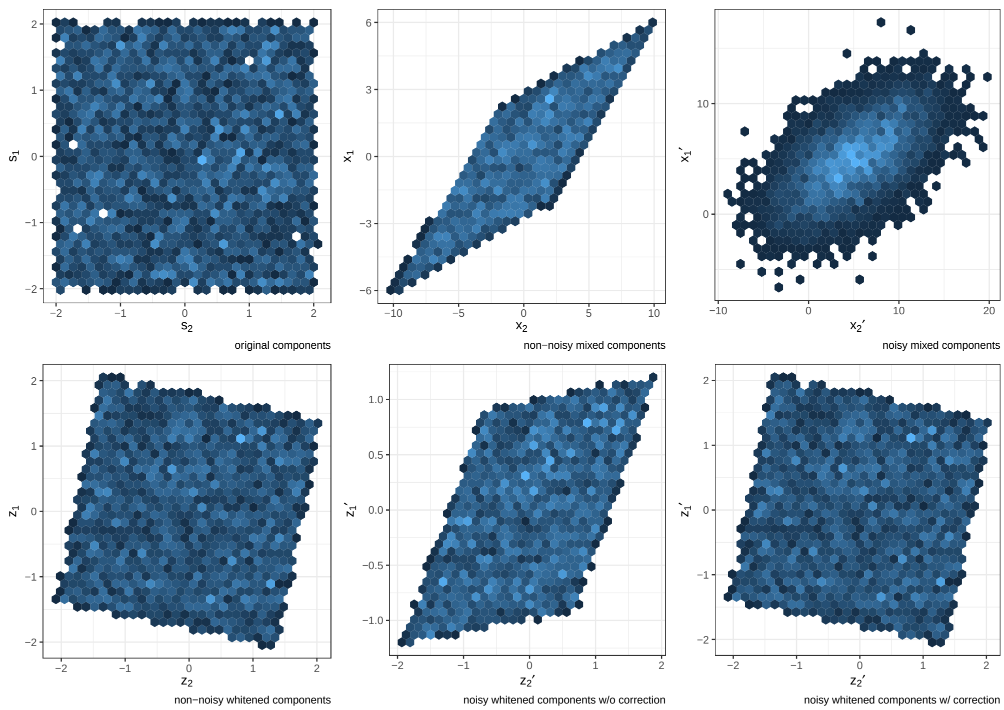

Figure 2: Illustration of whitening in non-noisy ICA setting and why whitening can be unfruitful in a noisy ICA setting.

setting. This shows that whitening can be implemented meaningfully in noisy settings, but it has the very strict requirement of knowing the noise distribution. In practical settings this is rarely known and it cannot be measured as the mixed components are not observed.

In conclusion, this example illustrates why whitening is not meaningful in an inherent noisy ICA setting. One can simply not guarantee a reduction in the search space and the step is thus unwarranted.

2.3 Retreating to a Safe Domain

Assume the ICA is done correctly and the computer outputs some vector $\mathbf{s}$[3]. The question now becomes: What does $\mathbf{s}$ signify and what can be inferred from it?

As it turns out there are a couple ambiguities that makes a *true recovery*[4] of the original components impossible.

First we cannot identify the order of the components given only the $x^{(i)}$'s, i.e. we can never know whether $\mathbf{A}^{-1}$ or $\mathbf{PA}^{-1}$, where $\mathbf{P}$ is some permutation matrix, is the true unmixing matrix since both A and s are unknown.

Further we cannot identify the true scaling of the original components either. For example consider the scaling of the mixing matrix $^1/_\alpha A$, where $\alpha \in \mathbb{R} \setminus \{0\}$. If we scaled the original components accordingly as $\alpha \cdot s_i$ we would obtain the same $x^{(i)}$'s. In general we can scale each

[3]This is of course a hurdle in and of itself.

[4]True is meant in the literal sense as being in (strict) accordance with reality.

	Method	Signal type	Allowed noise
1.	fastICA	non-Gaussian	none
2.	SOBI	fixed time-dependence	time-independent
3.	choiICA (var)	varying variance	none
4.	coroICA (var)	varying variance	group-wise stationary

Table 1: The 4 different ICA algorithms we will be studying. Here including the different assumptions they impose on the signal type and the possible noise. Table aggregated from [Pfister et al., 2019, p.5]

component with a non-zero scalar and obtain the same x_i if we multiply each column in the mixing matrix with the inverse. This also means that the variance of each component s_i is also impossible to recover as $\mathrm{var}(\alpha s_i) = \alpha^2 \mathrm{var}(s_i) \neq \mathrm{var}(s_i)$ in general. In summary, we cannot know which order the components are in or even their (specific) values. An illustration can be found in Appendix A.

Maybe surprisingly, this does not matter for the applications we are considering. First returning to the picture problem. The ambiguity of the order of components means that we cannot know which order the original pictures are in. The ambiguity of scaling means that the brightness of each picture is unidentifiable, but as we re-color the pictures after ICA nothing changes as their relative distance is preserved. Note the special case where the scaling is negative which inverts the color as seen in picture 1 and 2 in Figure 1.

Regarding EEG signal, it could be interesting in itself to know exactly where in the brain each signal come from. This has been a research field known as *Source Localization* since the early days of EEG signals in the 1950's, well before ICA became an area of research [Christoph and Michel, 2019]. This is however outside the scope of the present paper. Overall the ambiguities of the ICA are important concessions to be aware of but they will not stifle the specific questions we are asking.

2.4 The Algorithms We will in this section describe the different ICA algorithms. There are a lot of different algorithms in the ICA literature that deals with various aspects of the problem, let it be computation time or methodology for certain applications. It would therefore not be possible to compare all algorithms, the most we can hope for is to take some algorithms which can be deemed representative for the wealth of different algorithms in the literature. We will in this paper loosely follow the blueprint set out by [Pfister et al., 2019] and compare the 4 algorithms listed in Table 1.

2.4.1 fastICA The fastICA algorithm was introduced in [Hyvaärinen, 1999]. It tries to maximize negentropy by a transformation of the random variables. In fastICA a *non-Gaussianity* condition is assumed from the onset in order to guarantee identifiability (up to scaling and permutation).

The intuition behind fastICA and the somewhat peculiar condition of non-Gaussianity is the Central Limit Theorem (CLT). Since the observed components are a linear mixture of the original components they should be more Gaussian than the original components themselves due to the

CLT. Thus the signals could heuristically be unmixed by maximizing non-Gaussianity.

It should be noted that fastICA can be extended to be robust towards Gaussian noise, i.e.

$$\epsilon \sim \mathcal{N}(0, \mathbf{a}I_n), \quad \mathbf{a} \in \mathbb{R}^n.$$

2.4.2 SOBI The SOBI algorithm was first proposed by [Belouchrani et al., 1997] and differs from the fastICA in that it uses a *(weak) stationarity* assumption to approximate the unmixing matrix. More precisely it assumes the auto-covariance only depend on the time-lag τ not on the time t: [Nordhausen, 2014, p.143]

$$E(s_t s'_{t+\tau}) = \mathbf{D}_\tau \text{ is diagonal for all } \tau = 1, 2, \dots,.$$

In practice all lags up to a certain point will be chosen and the algorithm will then try to jointly diagonalize the corresponding auto-covariance matrices [Nordhausen, 2014]. The SOBI algorithm guarantees identification (up to scaling and permutation) for ICA's with time-independent noise.

2.4.3 choiICA & coroICA Both the choiICA and the coroICA algorithms use a *non-stationarity* assumption to approximate the unmixing matrix. Both algorithms exist in two main forms; a varying variance version (var)[5] and a time-dependence version (TD). For simplicity we focus on the varying variance version of the two algorithms in this paper. Both algorithms try to jointly diagonalize well-chosen blocks of covariances, including differences thereof in coroICA, to obtain the unmixing matrix. The main difference between the two algorithms is the allowed noise.

[Pfister et al., 2019] proposes the coroICA algorithm as the more robust algorithm as they prove identifiability of the unmixing matrix under group-wise stationary noise (cf. [Pfister et al., 2019, p.8]). The result stands in contrast to the choiICA (var) algorithm which does not guarantee identifiability for any type of noise. See further [Choi and Cichocki, 2000a, Choi and Cichocki, 2000b] and [Pfister et al., 2019].

Before leaving the discussion of different algorithms and their underlying assumptions, we feel compelled to repeat the common aphorism often attributed the the famous British statistician George Box: *All models are wrong, but some are useful*[6].

It could be hypothesized that ICA algorithms which accommodate more complex noise structures perform better than those that do not on noisy, complex data like the EEG data we consider in this paper. It goes without saying that an awareness of the different model assumptions is important, but we believe the final benchmark should always be usefulness, meaning that more refined models not necessarily imply better models in practice. Usefulness of course needs to contextualized which is done Section 3.2.

[5]Sometimes called the SD version, see e.g. [Nordhausen, 2014].

[6]The first instance of Box expressing this sentiment can be found in [E. P. Box, 1976].

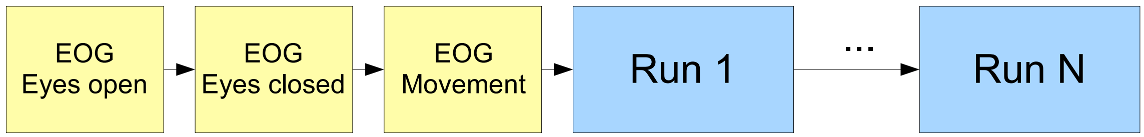

Figure 3: An illustration of the run setup. Each session is comprised of 6 runs. We discard the three first blocks of calibration. Figure taken from [Brunnder et al., 2008].

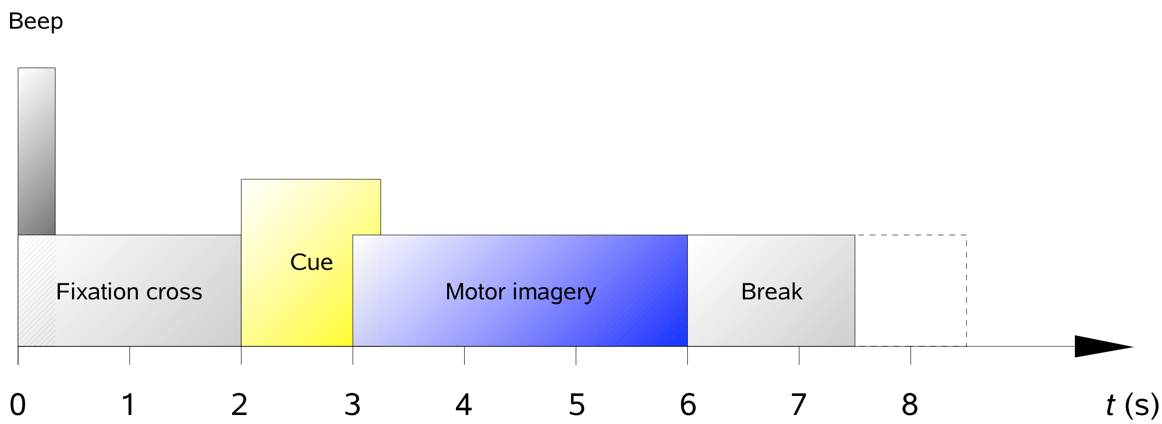

Figure 4: An illustration of the trial setup. Each run in Figure 3 is had 48 trials. We only use the data from second 3 to 6, i.e. the motor imagery box, in our analysis. The sampling rate was 250 Hz, meaning we have 3×22 data points for each trial. Figure taken from [Brunnder et al., 2008].

3 Methods & Results

In this section we describe our data, our methods and results for our analysis. All code can be found in this stable GitHub[7] repository.

3.1 Data Description The EEG data we consider in this paper was original used in the BCI IV case-competition (see [Tangermann et al., 2012] for the review article). Note here their objective was different than ours, it will therefore not be possible to compare our results directly.

The dataset consists of EEG data from 9 subjects performing 4 different motor imagery tasks. Each person recorded two sessions on different days. Each session consists of 6 runs where each run consists of 48 trials (where the 4 motor imagery tasks are evenly distributed), totaling 288 trials per session per person. Each recording had 22 EEG channels sampled at 250 Hz.

There has already been conducted some preprocessing on the data. A bandpass filter between 0.5 and 100 Hz and a 50 Hz notch filter have been applied. Furthermore an expert has marked all suspected artifacts, e.g. accidentally eye movement or muscle cramps, which we remove from the data from the onset. We have decided that this step does not warrant a complete removal of any

[7] https://github.com/Path-to-Exile/ICA-prject/tree/b66d3bf7fa5610f5b4380410f5bc6fd35380cc99

persons data, although there will be some discrepancy between the number of valid trials between some persons.

3.2 The Pipeline Study The computation required for our comparison analysis grows exponentially in the number of hyper-parameters in consideration. We therefore had to make some choices ourselves. We settled with 3 knobs each having two settings to turn, resulting in 8 different pipelines. Our metric for performance was prediction accuracy of which motor imagery task the subject was performing.

It could be fruitful to consider this performance criterion before continuing. Other performance criteria exist, e.g. the κ coefficient which [Tangermann et al., 2012] used in the BCI case competition. These and similar criteria, however, have one clear problem. They are not assessing how good the ICA method is in itself. They are only assessing the performance in conjunction with multiple other, possible just as important, components. Granted, it is hard to point to a performance assessing method without this fault as no one knows how true brain-signals should look like. This should not refrain anybody from doing data analysis on EEG data, but it should be kept in mind.

Our pipeline consisted of the following steps. The steps in **bold** were changed between pipelines.

1. Initial Preprocessing
2. **Projecting onto the orthogonal complement of the null component**
   - Yes
   - No
3. ICA method
   - 1. fastICA / 2. SOBI[8] / 3. choiICA (var) / 4. coroICA (var)
4. **Bandpass filter**
   - β band
   - α and β band
5. Feature Extraction
6. **Classification**
   - Quadratic Discriminant Analysis
   - Random Forest

We used Python for all our EEG pipelines. We used SciKit-learn [Pedregosa et al., 2011] for the implementation of the fastICA and for the classification algorithms used. We used the coroICA

[8] I was never able to resolve an issue regarding an unreasonably growing condition of number of unmixing matrices for the SOBI algorithm. I have still chosen to include my SOBI results, but the resulting accuracy score in each pipeline may be a bit lower than expected since the algorithm may have terminated early.

package [Pfister et al., 2019] for implementation of the SOBI, choiICA (var) and coroICA (var) algorithms. We used the SciPy package [Virtanen et al., 2020] for bandpass filtering. Discussion on the initial preprocessing and the ICA methods are already described in Section 2.4 and 3.1.

3.2.1 Common Average Reference Another preprocessing step we employed was projecting the data onto the orthogonal complement of the null component. Note that this amounts to doing a CAR (common average reference) which is a standard method to reduce spatial noise between electrodes [Christoph and Michel, 2019, p.176]. To ease the discussion of this, we refer to this simply as CAR which constitutes our first knob as we tested the ICA's performance with and without it.

3.2.2 Band-Pass filtering Before feature extraction we bandpass filtered our data. Unfortunately EEG bands is a point of contention in the literature as no standard exists[9]. Different authors will define the same EEG bands in slightly different ranges and in turn their bandpass filtering will be different. In this paper we have chosen to define the frequency range of the α band as $]8-13]$ Hz and the frequency range for the β band as the $]13-30[$ Hz, which is in line with the most common cut-offs found in the literature according to [Newson and Thiagarajan, 2019].

How we implemented the band-pass filtering is the second knob in our pipelines. a) We bandpass filtered our data in the β band and the extracted the band-power features as the logarithmic variance. b) We employed two bandpass filters one in the β range and one in the α range. We then extracted band-power features as the logarithmic variance for both of the filters.

3.2.3 Classifier Our third knob was choice of classifier. We used an Quadratic Discriminant Analysis classifier with a shrinkage parameter of 0.1 and an ordinary Random Forest classifier. We then assessed accuracy against the known true values.

For our comparison experiment we trained on a random sample of 4 subjects and tested against the remaining 5 subjects. We performed each pipeline 20 times sampling a random subset of 4 subjects. We summarize our result in Figure 5 and Table 2.

We calculated *Beat in percentage* for each combination of ICA methods as the percentage of times the first ICA method resulted in better performed than the second ICA wrt. to subset of subjects and pipeline choice. *Difference in percentage point* was similarly calculated as average percentage point difference in performance wrt. specific subset of subjects and pipeline choice.

Using `R` we fitted a linear model using the pipeline choices as main effects and accuracy as response which can be seen in Table 3. We corroborate on the meaning and the interpretation (especially limitations thereof) in Section 4.

[9] cf. [Newson and Thiagarajan, 2019] for a discussion on this.

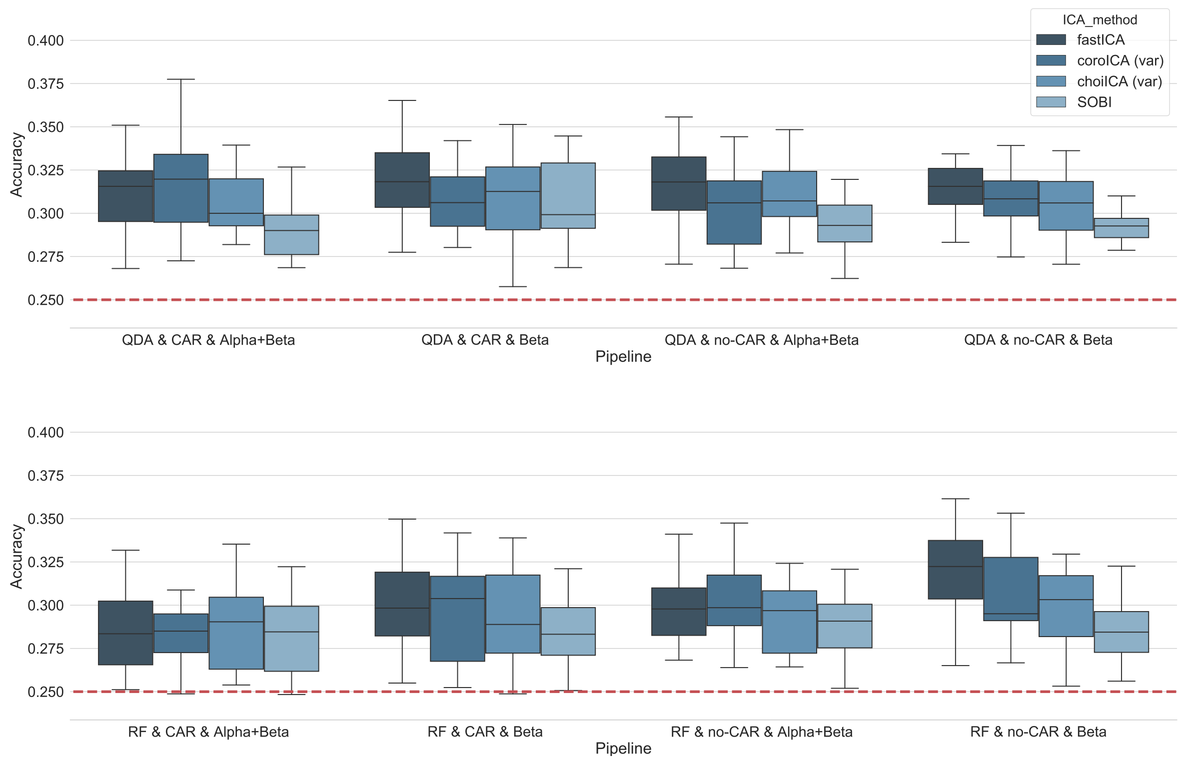

Figure 5: Results of all eight pipelines sorted by ICA method used. alpha/beta means we used both an α and a β bandpass filter. The dashed red line represent an accuracy of 25% which correspond to a random guess.

	Pipeline	Accuracy
1	QDA, CAR, alpha+beta	0.306
2	QDA, CAR, beta	0.313
3	QDA, no-CAR, alpha+beta	0.306
4	QDA, no-CAR, beta	0.305
5	RF, CAR, alpha+beta	0.285
6	RF, CAR, beta	0.294
7	RF, no-CAR, alpha+beta	0.296
8	RF, no-CAR, beta	0.302

ICA vs.	Beat in pct.	Dif. in pp.
fastICA vs. SOBI (var)	0.756	1.69
coroICA vs. SOBI (var)	0.731	1.29
choiICA vs. SOBI (var)	0.706	0.934
fastICA vs. choiICA (var)	0.700	0.760
fastICA vs. coroICA (var)	0.619	0.403
coroICA vs. choiICA (var)	0.550	0.357

Table 2: To the left we see each pipeline with their mean accuracy score over all ICA methods. To the right we see a sorted comparison of accuracy scores between each ICA method for each subset of subject over all pipelines. We emphasize again that we only compared the ICA methods under the exact same conditions.

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.2966	0.0025	119.68	0.0000
bandbeta	0.0052	0.0019	2.80	0.0053
clfRF	-0.0134	0.0019	-7.18	0.0000
ICA_methodchoiICA (var)	0.0093	0.0026	3.52	0.0005
ICA_methodcoroICA (var)	0.0129	0.0026	4.87	0.0000
ICA_methodfastICA	0.0169	0.0026	6.39	0.0000
CAR	-0.0028	0.0019	-1.52	0.1296

Table 3: `R` output for our fitted LM.

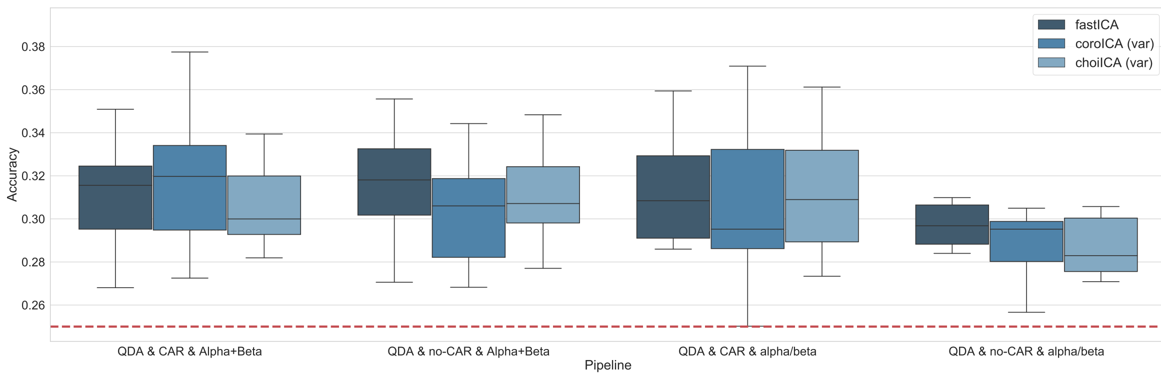

Figure 6: Comparison between pipelines where we used two separate bandpass filters, one in the α range and one in the β range, and pipelines where we used one bandpass filter in the $\alpha + \beta$ range, i.e. $]8, 30[$ Hz range.

3.3 Further Experiments We also compared the QDA pipelines where we used an α and a β bandpass filter from before to pipelines where we used a wider bandpass filter encompassing both the α and β range. We used the same methodology as before with the exception that we dropped the SOBI algorithm for our new pipeline comparison. Our results can be seen in Figure 6. The difference in performance between using two filters vs. using one larger filter over both pipelines (with and without CAR) was 0.790 percentage points in favor of using one larger filter. Note however we re-sampled the subsets for the two new pipeline.

We choose to pursue a larger analysis on the pipeline without CAR, β-band filtering, QDA classifier and only considering FastICA, coroICA and choiICA. We trained on 1 through 9 subjects, each time taking a random sample with replacement, and tested against the remaining subjects. We performed this 20 times for each number of training subjects. Our results are summarized in Figure 7 and Table 4 in Appendix B.

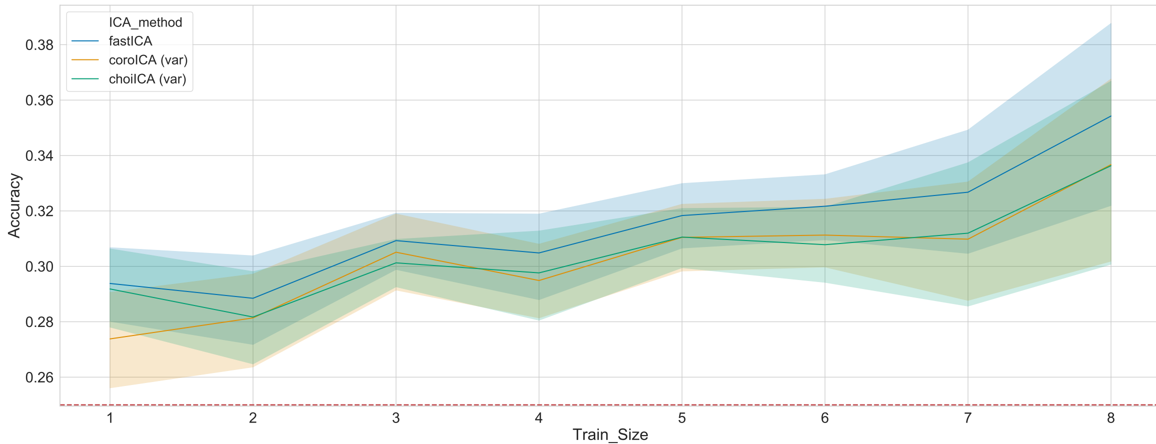

Figure 7: The accuracy of each ICA method for each number of subjects trained on.

4 Discussion

In this section we analyze and discuss our results. Throughout this section we write phrases like "*ICA method A performed better than ICA method B*". This is slight misnomer as the ICA methods can only be evaluated as an (unsupervised) part of each pipeline. Statements like these should therefore be understood as "*In pipeline* 1 *(or across some pipelines), using ICA method A resulted in better performance than using ICA method B*". We prefer the former as it is less tedious to read and write.

4.1 Pipeline Discussion We cannot unequivocally pronounce one best performing pipeline from our experiment. However, our analysis does suggest that pipelines with a QDA classifier performs better than pipelines with a RF classifier, although it should be mentioned that the pipeline with RF, no CAR and β-band had the highest mean accuracy score using fastICA of all pipelines.

4.1.1 Underachieving SOBI SOBI performed worse than the three other ICA algorithms for all pipelines, with the biggest difference being between fastICA and SOBI. fastICA performed better than SOBI about 76% percent of the time when comparing the two ICA's performance on the same subsets. The mean difference in accuracy was 1.7 percentage point.

SOBI's relatively bad performance may in part be because of the issue with the implementation mentioned earlier, but SOBI also performed worse than the other three algorithms across all training set sizes in [Pfister et al., 2019]. We therefore believe it is a fair assessment to say SOBI is the worst of the four ICA methods in our pipelines.

4.1.2 Many ICA's Lead to Rome The three other algorithms relative performance was closer. fastICA was slightly above both choiICA and coroICA outperforming them in respectively 70% and

62% of the cases. choiICA and coroICA performance was very close to each other and the best performing algorithm changed between pipelines, most likely due to the relative low sample size of 20. Although when comparing the two algorithms on a subset basis coroICA had a slight edge over choiICA as it performed better in 55% of the cases with a mean accuracy difference of 0.36 percentage points.

4.1.3 Knob Discussion From the plots and accuracy values alone it is hard to decide whether or not we gained predictive power using both the α- and β-band instead of just the β-band. But we can conclude that using both the α- and β-band did not in general outperform only using the β-band, indicating that more features does not necessarily result in better performance.

The picture becomes muddier when comparing the use of two bands to the use of one wider band encompassing both the α- and β-range in Figure 6. An experimental setup with one wider bandpass filter resembles the one seen in [Pfister et al., 2019] and multiple teams who worked on the present dataset in the BCI competition [Brunnder et al., 2008] [10].

The accuracy scores of the two pipelines with the wide band were in general higher than those with the two smaller bands with an average mean difference of 0.790 percentage points. Despite this, it could be interesting in the future to investigate whether a smaller band, which possibly blocks out more unwanted noise while retaining the most important brainwave signals, can lead to better performance. In addition we would also encourage a standardization of EEG bands, as discussed in [Newson and Thiagarajan, 2019], in order to have a better basis for reference.

Further we showed no link between using CAR and predictive power. The accuracy scores are very similar when comparing the same pipelines with and without the use of CAR. This falls in line with comments in [Yao et al., 2019] who questions the viability of average reference techniques in modern EEG data analysis.

The best performing pipeline when aggregating the four ICA methods was **QDA, CAR and β-band** with an accuracy score of 0.313. The next tier of pipelines (in descending order according by accuracy score) were **QDA, CAR, α-band and β-band**, **QDA, no CAR, α-band and β-band**, **QDA, no CAR and β-band** and **RF, no CAR, β-band** with accuracy scores between $0.302 - 0.306$. This corroborates on that the QDA classifier results in better performance than the RF classifier as the four pipelines with QDA all had higher mean accuracy scores than those with RF classifier.

4.1.4 The Linear Model We also made a linear model with the different pipeline choices and different ICA methods as main effects. Before discussion the results, the limitations of such a model needs to be clear. *Independence between observations is not fulfilled* since the individual subjects overlap between subset samples. We should therefore be vary of how much weight we put on the

[10]See here for a short overview of the different pipelines (`http://www.bbci.de/competition/iv/results/index.html`). The pipelines are under dataset 2a

specific estimates and p-values. The most we can hope for is that our linear model can somehow illuminate our present problem.

Our linear model agrees on most of the trends we have already discussed. The model agrees in that SOBI is the worst ICA method, that the RF classifier is worse than the QDA and that CAR is not particularly important for high accuracy scores. Furthermore the LM suggests that one obtains better accuracy by only considering the β-band which is interesting since it would mean the information obtained from the α-band is actually detrimental to the classifiers ability to correctly guess the given motor imagery action.

4.2 Multi Subset Size Discussion We choose to make our multi subset size analysis on the pipeline with **QDA, no CAR and $\boldsymbol{\beta}$-band**. From our pipeline analysis we believe that we could have used any of the QDA pipelines and obtained a very similar plot. However computational limitations forced us to only use one pipeline. From the earlier discussion we also choose not to use SOBI as it seems to perform the worst of the four algorithms. In general we can conclude that more training material results in better accuracy scores. Again fastICA performed the best, with choiICA and coroICA yielding very similar performances.

4.2.1 Comparison to Pfister et al. We have lower accuracy scores across all subset sizes when comparing our results with the similar setup found in [Pfister et al., 2019, p.42] . One of the primary reasons for their superior scores may be, that they resampled by bootstrapping the training and test trials 50 times for each particular split of data to get a more stable estimate of the achieved classification accuracy.

Some of our obtained accuracy scores are not, or only very little, above random guess. Resampling could mitigate this as our analysis would be less prone to outliers. It would also relieve the problem of over-representation of outliers. However, due to computational constraints we were not able to perform this type of resampling.

Other factors which could explain the the discrepancy between accuracy scores is the use of a different classification algorithm. We used a standard of the shelves classifier, whereas they used a more sophisticated classifier (cf. [Pfister et al., 2019, p.29] for details).

Further coroICA (var) performed substantially better in [Pfister et al., 2019] outperforming fastICA, and the rest of the ICA methods, for the biggest subset sizes. We were not able to reproduce these results. More testing, e.g. by looking at different parameter settings or runtime, is needed to determine whether it is our results or [Pfister et al., 2019] results that need to be adjusted.

4.3 Conclusion Using fastICA resulted in the best pipeline performance in both our pipeline study and multi subset size study. coroICA and choiICA performed similarly while SOBI consistently was the worst performing ICA method. We were however never able to get any pipeline configuration consistently above an accuracy percentage in the low 30's.

We conclude from our pipeline study that the most important factors for obtaining a high accuracy in ones pipeline is choosing a suitable classifier and not using a bad ICA method, i.e. SOBI in this paper. Conversely, we can also conclude that a lot of different pipeline choices leads to very similar results. Due to the size of our study our results can only be deemed preliminary, but if our conclusions should hold in more generality it would have two positive consequences:

1. There exist no clear bias inducing hyper-parameter choices.
2. The apparent lack of standardization in the literature (cf. [Newson and Thiagarajan, 2019]) does not mean that the obtained results of the individual papers are misguiding. Other experts would more or less have obtained the same results despite their different choices in pipeline setup.

The multi subset size study further echoes our earlier discussion about the ICA methods only being one part of a larger pipeline with other equally important components. [Pfister et al., 2019] had better performance across all subset sizes despite the use of the same ICA methods, pointing to other factors, like more advanced sampling techniques and better classification algorithms, can have an even bigger role on performance than the specific ICA method used.

A Ambiguities in the 2×2 case

We can exemplify these ambiguities in the 2×2 case. The ambiguities extend to arbitrary dimensions, but for notational ease we will only show it for 2 dimensions. One can easily extend the presented ideas to the general n dimensional case.

We define a 2 dimensional permutation matrix $\mathbf{P}$ as

$$\mathbf{P} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

We see that given an observed $\mathbf{x}$ we can change the order of each row in $\mathbf{A}$ if we also change the order of the elements in $\mathbf{s}$:

$$\begin{aligned}
\mathbf{x} = \mathbf{A}\mathbf{s} = \mathbf{A}\mathbf{P}\mathbf{P}^{-1}\mathbf{s} &= \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} s_1 \\ s_2 \end{pmatrix} \\
&= \underbrace{\begin{pmatrix} a_{12} & a_{11} \\ a_{22} & a_{21} \end{pmatrix}}_{\mathbf{A}'} \underbrace{\begin{pmatrix} s_2 \\ s_1 \end{pmatrix}}_{\mathbf{s}'}.
\end{aligned}$$

Since both $\mathbf{A}$ and $\mathbf{s}$ are unknown we can never know whether $\mathbf{A}$ or $\mathbf{A}'$ is the true mixing matrix and in turn whether $\mathbf{s}$ or $\mathbf{s}'$ is the true order of original components.

Let $\alpha, \beta \in \mathbb{R} \setminus \{0\}$. From the following we see that it is possible to extend the original components with any given non-zero scalar if we correct the mixing matrix accordingly:

$$\mathbf{x} = \mathbf{As} = \begin{pmatrix} a_{11}s_1 & a_{12}s_2 \\ a_{21}s_1 & a_{22}s_2 \end{pmatrix} = \begin{pmatrix} \frac{1}{\alpha}a_{11} \cdot \alpha s_1 & \frac{1}{\beta}a_{12} \cdot \beta s_2 \\ \frac{1}{\alpha}a_{21} \cdot \alpha s_1 & \cdot\frac{1}{\beta}a_{22} \cdot \beta s_2 \end{pmatrix} = \begin{pmatrix} \frac{1}{\alpha}a_{11} & \frac{1}{\beta}a_{12} \\ \frac{1}{\alpha}a_{21} & \frac{1}{\beta}a_{22} \end{pmatrix} \begin{pmatrix} \alpha s_1 \\ \beta s_2 \end{pmatrix}$$

$$= \underbrace{\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} \frac{1}{\alpha} \\ \frac{1}{\beta} \end{pmatrix}}_{\mathbf{A}'} \underbrace{\begin{pmatrix} \alpha s_1 \\ \beta s_2 \end{pmatrix}}_{\mathbf{s}'}.$$

Again, since both $\mathbf{A}$ and $\mathbf{s}$ are unknown we can never know whether $\mathbf{A}$ or $\mathbf{A}'$ is the true mixing matrix and further whether $\mathbf{s}$ or $\mathbf{s}'$ vector is the true vector of the original components.

B Multi Subset Experiment Table

Training set size	fastICA	choiICA (var)	coroICA (var)	Aggregated
1	0.294	0.292	0.274	0.286
2	0.288	0.282	0.281	0.284
3	0.309	0.301	0.305	0.305
4	0.305	0.298	0.295	0.299
5	0.318	0.311	0.310	0.313
6	0.322	0.308	0.311	0.314
7	0.327	0.312	0.310	0.316
8	0.354	0.336	0.337	0.342

Table 4: Accuracy aggregated according to Training set size and ICA method. We used the pipeline consisting of no CAR, only β-band filtering and QDA classifier.

References

[Belouchrani et al., 1997] Belouchrani, A., Abed-Meraim, K., Cardoso, J., and Moulines, E. (1997). A blind source separation technique using second-order statistics. *IEEE Transactions on Signal Processing*, 45:434–444.

[Brunnder et al., 2008] Brunnder, C., Leeb, R., Müller-Putz, G. R., Schlögl, A., and Pfurtscheller, G. (2008). Bci competition 2008 – graz data set a. http://www.bbci.de/competition/iv/desc_2a.pdf.

[Chakravartty, 2017] Chakravartty, A. (2017). Scientific Realism. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University.

[Choi and Cichocki, 2000a] Choi, S. and Cichocki, A. (2000a). Blind separation of nonstationary and temporally correlated sources from noisy mixtures. *Proceedings of the 2000 IEEE Signal Processing Society Workshop*, page 405–414.

[Choi and Cichocki, 2000b] Choi, S. and Cichocki, A. (2000b). Blind separation of nonstationary sources in noisy mixtures. *Proceedings of the 2000 IEEE Signal Processing Society Workshop*, 36(9):848–849.

[Christoph and Michel, 2019] Christoph, M. and Michel, B. H. (2019). *EEG source Localisation*, pages 85–101. Elsevier Science, Netherlands.

[E. P. Box, 1976] E. P. Box, G. (1976). Science and Statistics. *Journal of the American Statistical Association*, 71(356):791–799.

[Hyvaärinen, 1999] Hyvaärinen, A. (1999). Fast and Robust Fixed-Point Algorithms for Independent Component Analysis. *IEEE Trans. on Neural Networks*, 10(3):626–634.

[Hyvärinen et al., 2001] Hyvärinen, A., Juha., K., and Oja, E. (2001). *Independent Component Analysis*. Wiley, New York, USA.

[Hyvärinen and Oja, 2000] Hyvärinen, A. and Oja, E. (2000). Independent Component Analysis: Algorithms and Applications. https://www.cs.helsinki.fi/u/ahyvarin/papers/NN00new.pdf.

[Newson and Thiagarajan, 2019] Newson, J. J. and Thiagarajan, T. C. (2019). EEG Frequency Bands in Psychiatric Disorders: A Review of Resting State Studies. *Frontiers in Human Neuroscience*, 12:1–24.

[Nordhausen, 2014] Nordhausen, K. (2014). On robustifying some second order blind source separation methods for nonstationary time series.

[Pedregosa et al., 2011] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.

[Pfister et al., 2019] Pfister, N., Weichwald, S., Bühlmann, P., and Schölkopf, B. (2019). Robustifying Independent Component Analysis by adjusting for Group-Wise Stationary Noise. *Journal of Machine Learning Research*, 20:1–50.

[Tangermann et al., 2012] Tangermann, M., Müller, K.-R., Aertsen, A., Birbaumer, N., Braun, C., Brunner, C., Leeb, R., Mehring, C., Miller, K., Mueller-Putz, G., Nolte, G., Pfurtscheller, G., Preissl, H., Schalk, G., Schlögl, A., Vidaurre, C., Waldert, S., and Blankertz, B. (2012). Review of the BCI Competition iv. *Frontiers in Neuroscience*, 6.

[Virtanen et al., 2020] Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Jarrod Millman, K., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., Carey, C., Polat, İ., Feng, Y., Moore, E. W., Vand erPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E. A., Harris, C. R., Archibald, A. M., Ribeiro, A. H., Pedregosa, F., van Mulbregt, P., and Contributors, S. . . (2020). SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272.

[Yao et al., 2019] Yao, D., Qin, Y., Hu, S., Dong, L., Vega, M. L. B., and Sosa, P. A. V. (2019). Which reference should we use for EEG and ERP practice? *Brain topography*, pages 1–20.