SPATIAL FILTERS FOR AUDITORY EVOKED POTENTIALS TRANSFER BETWEEN DIFFERENT EXPERIMENTAL CONDITIONS

Jan Sosulski[1], Michael Tangermann[1,2]

[1]Brain State Decoding Lab, Cluster of Excellence BrainLinks-BrainTools, Dept. Computer Science, University of Freiburg, Freiburg, Germany
[2]Autonomous Intelligent Systems Lab, Dept. Computer Science, University of Freiburg, Freiburg, Germany

E-mail: { jan.sosulski | michael.tangermann } @blbt.uni-freiburg.de

ABSTRACT: To interact with the brain in closed-loop applications despite low signal-to-noise ratios, brain-computer interfaces can make use of data-driven linear spatial filtering approaches to improve the single-trial classification performance. While the transfer of spatial filters between users and within multiple sessions of the same user under the same experimental paradigm is feasible to some extent, it is unclear, whether changing experimental conditions affects this transferability.

To investigate this question with regards to spatial filters, we evoke event-related potentials by an auditory oddball paradigm under various stimulus onset asynchronies (SOAs). Using four similarity and distance measures, we analyze the between and within subject transferability of xDAWN filters.

We found that the used measures reflect the differences of spatial filters between subjects. Within the same subject, the measures indicate similarity of the spatial filters for almost all SOA conditions. We conclude, that our proposed measures can be used to give indications under which circumstances spatial filters can be transferred.

INTRODUCTION

Brain-computer interfaces (BCIs) are systems which use machine learning to interpret the brain signals of its users in single trial in order to obtain information and possibly act upon this information. Popular applications for BCI are the control of wheelchairs [1], spelling applications for patients [2, 3], stroke rehabilitation [4] and non-clinical applications [5, 6]. One non-invasive way of obtaining the brain signals is to use electroencephalogram (EEG) recordings which are obtained via electrodes placed on the scalp of user.

In order for the brain signals to be recorded by the EEG, they have to travel through the brain, the skull and the skin of the subject. Having low signal amplitudes, they are easily masked by non-neural signal sources as the sensitive electrodes collect all sorts of electric activity, such as power line frequency, muscle activation or currents induced by electrode movements. These interfering electric activities are called artifacts. As a result, the obtained neural signals have generally a poor signal quality, i.e., a bad signal-to-noise ratio. In BCI, one is usually interested only in a specific type of brain signal while background EEG activity and artifacts are to be ignored. Frequently used techniques to deal with these problems are frequency filtering [7], spatial filtering [8, 9] and artifact removal [10]. Applying these methods is oftentimes necessary to extract useful information from the EEG signal.

The aforementioned spatial filtering methods try to find weights on electrodes that minimize the influence of noise while maximizing the brain signal of interest. General purpose approaches such as Laplacian filtering [11] can usually be applied without special considerations. However, subject-specific data-driven spatial filtering methods, e.g., common spatial patterns [8] or SPoC [12] for oscillatory brain signals or xDAWN [9] for event-related potentials (ERPs), usually outperform these general approaches. The drawback of subject-specific methods is that the filters have to be trained on data from the subject, therefore requiring additional training data.

One experiment to elicit auditory ERPs is the simple two-class auditory oddball paradigm, where the subject is asked to attend to rare high-pitched target tones and to ignore frequent low-pitched non-target tones. While each tone stimulus elicits an ERP response, they have different spatio-temporal characteristics depending on the class (i.e., target or non-target) of the played tone. Additionally, the waveform of these ERPs can look differently depending on other stimulation parameters. In this work, we investigate the influence of stimulation onset asynchrony (SOA), which denotes the time between the onset of two successive tone stimuli. The SOAs we examine are in the range of 60 ms to 600 ms. The so-called P300 response is a positive potential occurring approximately 300 ms after a target stimulus over central electrodes. This ERP component can be evoked by oddball tasks. Its delay relative to the stimulus onset has been reported to vary with task difficulty [13], among others. Thus the P300 rise may occur earlier in simple auditory oddball tasks. Besides the P300, other modality-dependent and task-dependent ERP components can be elicited, which have shorter or longer delays compared to

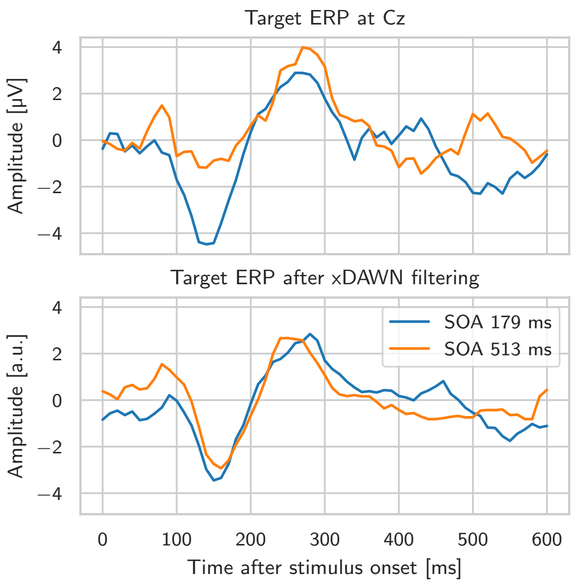

Figure 1: Averaged ERP waveforms elicited by auditory target stimuli delivered at time t=0 ms for two SOAs. Top: waveforms observed at channel *Cz* (no spatial filter was applied). Bottom: waveform observed after applying an xDAWN spatial filter.

the P300. For multiple stimuli presented at short SOAs, it thus is known, that elicited ERP responses do overlap. The used xDAWN method is specifically suited to reduce this overlap in the filtered ERP responses.

Figure 1 shows examples, how changing the SOA parameter can influence the averaged waveform of the ERP responses recorded at the *Cz* electrode. However, when looking at the same signals filtered using an xDAWN filter, the average ERP responses look more similar. Please note, that in this example one xDAWN filter was determined specifically for each SOA condition based on training data.

Finding solutions to reduce the amount of necessary training data is beneficial in BCI for many reasons. Transfer learning of spatial filters and full classifiers between repeated sessions of one user or even between users has been investigated in this context [14]. For changing experimental conditions, however, it is unclear, if such transfer is feasible. Thus we will focus on the question, how well spatial filters trained under a specific SOA condition can be transferred to a different SOA condition, but within the same user.

To investigate this, we will first propose four measures that quantify transferability. Finally, we will apply these measures to within subject transfers of xDAWN filters obtained under different SOA conditions in an auditory oddball experiment.

MATERIALS AND METHODS

Data-driven spatial filtering algorithms learn a weight matrix W with $N_w \times N_c$ dimensions that is applied to the $N_c \times N_t$-dimensional EEG signal X with N_c channels and N_t number of time samples. The filtered signal

$$S = WX \tag{1}$$

retains $N_w \times N_t$ dimensions, while the number of filters N_w is usually much smaller than the original number of channels N_c. In this work we focus on a single individual spatial filter $\mathbf{w}$ only, such that we retain a weight vector instead of a matrix, i.e., $N_w = 1$. The physiological interpretation of spatial filter weights is not straightforward. As described by Haufe et al. [15], one usually calculates the activation pattern of a filtered signal to support the interpretation. This activation pattern A can be calculated using the channel covariance matrices of the original EEG signal Σ_x and the filtered EEG signal Σ_s

$$A = \Sigma_x W \Sigma_s^{-1}. \tag{2}$$

The covariance matrices are calculated for a time interval of interest. As we only consider a single spatial filter $\mathbf{w}$, we also look at only a single spatial pattern $\mathbf{a}$.

We use the xDAWN algorithm to find spatial filters $\mathbf{w}$ that enhance target ERPs elicited in an auditory oddball paradigm. The xDAWN spatial filter is obtained by solving a generalized eigenvalue problem. Therefore, the resulting spatial filters do not always have the correct—in the sense that the P300 ERP is actually positive—sign and have no longer interpretable amplitudes. We usually chose the spatial filter on the first rank, unless visual inspection suggested that the filters on second or third rank represented the P300 ERP better.

After receiving an ethics vote from the local ethics committee in Freiburg, and after obtaining written informed consent, EEG data from 13 subjects (six female, seven male, mean age: 24.9 years, standard deviations: 5.1 years) was recorded. Within a single session 20 trials (split in four blocks with five trials each) of an auditory oddball paradigm were conducted for three to five different SOA conditions, where one trial consisted of 15 target and 75 non-target stimulus presentations. In total this yields 20 trials for each SOA condition. The order of the stimuli was pseudo-randomized such that at least two non-target stimuli were presented between any two target stimuli. The recorded EEG data was preprocessed for offline analysis by zero phase band-pass filtering in a frequency band from 1.5 Hz to 40 Hz. While this frequency range is atypical for traditional ERP analysis, the usage of very short SOAs down to 60 ms required this higher-frequent upper limit of the low-pass. After frequency filtering, the signal was downsampled from 1000 Hz to 100 Hz. The signal was then epoched from −200 ms to 1000 ms around a stimulus. Each epoch was baseline corrected relative to the interval of −200 ms to 0 ms. We chose not to perform any artifact removal to keep the amount of data comparable between all subjects and to estimate how robust the xDAWN filter is under the influence of potential artifacts. For each subject and SOA condition we trained an individual xDAWN filter in the interval of 0 ms to 600 ms as suggested in [9]. Additionally, we trained an xDAWN filter on a mixture data set containing trials from each SOA condition such that its size was equal to the number of trials for the individual SOAs. As one xDAWN filter is obtained per SOA and one for the mixed data set, we finally obtain four to six spatial filters per subject. This yields a total of 66 different xDAWN spatial filters.

On this data we applied several similarity and distance measures that should allow to determine whether a spatial filter can be transferred to other conditions or subjects.

Comparing spatial filter weights:

In order to quantify commonalities or differences between two spatial filters, we can employ a similarity measure. The spatial filters determined by xDAWN are represented as weight vectors that unfortunately are determined only up to sign and amplitude. However, this representation still allows us to use the cosine similarity measure, i.e., the cosine of the angle θ between two weight vectors $\mathbf{w}_1$ and $\mathbf{w}_2$:

$$\operatorname{cossim}(\mathbf{w}_1, \mathbf{w}_2) = \cos(\theta) = \frac{\mathbf{w}_1 \cdot \mathbf{w}_2}{||\mathbf{w}_1|| \cdot ||\mathbf{w}_2||} \quad (3)$$

Angles between weight vectors has been proposed for comparing spatial filters before [16, 17]. A cosine similarity of 1 indicates collinearity of the vectors, while 0 indicates orthogonality. Negative values indicate that the vectors point to opposite half spaces. Please note, that the cosine similarity disregards spatial relations between electrodes.

Applied to our setup we expect smaller angles between xDAWN filters within one subject, i.e., $\cos(\theta) \to 1$. For analysis, we calculate pairwise cosine similarities of all possible pairs within the 66 xDAWN filters obtained over subjects and SOA conditions.

Comparing elicited patterns:

Aside from similarity of the actual spatial filters, the similarities of the elicited patterns may be informative. Patterns tend to be more homogeneous than spatial filters and the interesting information is the distribution and location of the pattern on the scalp. Therefore, instead of using the location invariant cosine similarity to compare patterns, we utilize the optimal transport (OT) distance [18] (also known as earth mover's distance or Wasserstein distance). This OT distance tries to find the minimal transport plan or flow F that transforms a distribution A into distribution B. In our case A and B are the elicited spatial patterns.

For our data set we expect smaller OT distances when comparing two patterns obtained for different SOAs within one subject than when comparing two patterns for different subjects.

Comparing ERP waveforms:

We also need to quantify differences between ERP waveforms, that result from spatially filtered EEG signals. As amplitude values generally are not comparable anymore after the application of xDAWN filters, we first rescale each filtered waveform to a standard deviation of one.

To lower the influence of noise prior to applying a waveform-specific distance metric, we reduce the temporal resolution of the waveforms to six predefined time

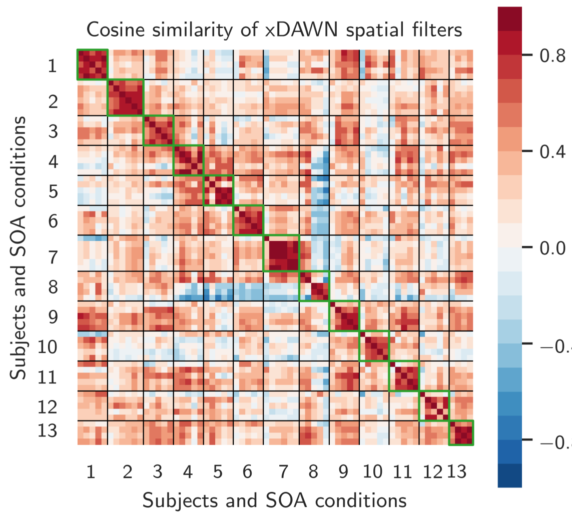

Figure 2: Matrix of all pairwise cosine similarities between xDAWN filters trained for each subject and SOA condition. The thick black lines separate subjects from each other. Note that the green-bordered blocks on the main diagonal correspond to within subject filter similarities under different SOA conditions.

intervals by averaging over multiple samples. Furthermore, we average the waveform features over trials of the same condition and subject. Thus, we obtain six amplitude features per SOA and subject.

Then the difference between the amplitude features of two filtered ERP waveforms $\mathbf{s}_1$ and $\mathbf{s}_2$ is quantified by using the Euclidean distance.

Classification performance:

We have limited our analysis to a single spatial filter per SOA condition and subject. However, if a spatial filter $\mathbf{w}$ captures the relevant features of the ERP responses evoked by the experimental paradigm, then applying it to a data set X should nevertheless lead to a target/non-target classification performance well above chance level. We will make use of a linear discriminant analysis classifier with covariance shrinkage regularization (rLDA) to estimate the class discriminative information provided by a filter, given by the area under receiver-operator characteristic curve (AUC). As a single xDAWN component $\mathbf{w}$ may not capture the full ERP information, we generally must expect a slight performance decrease by applying a single spatially filter compared to using the high-dimensional unfiltered signal [9]. To quantify this performance drop, we trained an rLDA classifier on each data set, first without applying any spatial filtering and then (consecutively) using each of the derived 66 xDAWN spatial filters. Utilizing a 4-fold chronological cross-validation scheme we can report the estimated performance loss $\Delta\text{AUC} = \text{AUC}_{\text{filt}} - \text{AUC}_{\text{nofilt}}$ for each spatial filter when transferred to each data set.

Successful transfer of filters should be expressed by an only small drop in performance.

RESULTS

Table 1: List of SOA conditions (in ms) used for each subject.

Subject	**SOAs in ms**				
1	60	175	200	460	—
2	60	111	200	218	446
3	60	123	175	380	—
4	60	193	286	506	—
5	60	208	226	508	—
6	60	175	296	474	—
7	60	226	235	518	596
8	60	177	256	446	—
9	60	180	400	447	—
10	60	220	325	402	—
11	60	179	194	513	—
12	60	193	220	600	—
13	60	211	518	—	—

First, we utilize the cosine similarity to investigate the similarity between xDAWN filters. In Figure 2, the cosine similarities between all spatial filters calculated for each subject and SOA condition are visualized. Black lines indicate the boundary between subjects. Within one subject, the SOA conditions on which individual filters were trained are in ascending order from top to bottom / left to right. The concrete SOAs for each subject are shown in Table 1. The last filter of each subject was trained on a mixed data set of all SOAs with an equivalent number of trials.

Intense red entries in the diagonal blocks indicate, that the similarities between spatial filters obtained from within a single subject are generally high. Specifically, they are considerably larger than filter similarities between subjects as indicated by the off-diagonal blocks. Within subjects, the fastest SOA of 60 ms seems to take a special role: its spatial filters are substantially different from the ones calculated under slower SOA conditions for most subjects. Additionally, for subjects 4, 5, and 12, there exist some SOA conditions which do not produce a spatial filter comparable to the rest. Interestingly, subject 8 produces spatial filters that are specific for the two fast SOAs and different spatial filters specific for the two slower SOAs. Additionally, the last row/column of each subject generally shows consistent similarities to other filters for this subject (data not shown). This observation is in line with our expectations, as this filter had been trained on the mixed data set consisting of all SOAs of the subject.

The histograms in Figure 3 provide a closer look at the within subject vs. between subject cosine similarities of the xDAWN filters. As expected, we can observe higher similarities of filters within a subject (blue color, median similarity 0.695) than across subjects (orange color, median similarity of 0.218). Introspection into the data revealed that values close to zero cosine similarity were caused mostly by filters of the very fast 60 ms SOA.

The optimal transport distance between patterns provides a novel view upon the similarity of xDAWN components. The histogram in Figure 4 shows, that the differences of the distance of patterns between and within subjects are not as emphasized as for cosine similarities which acts on

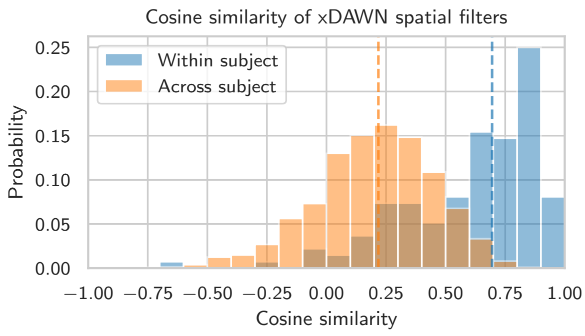

Figure 3: Distribution of the similarities between spatial filters within and across subjects. Dashed lines indicate median values of 0.218 (between subjects) and 0.695 (within subjects).

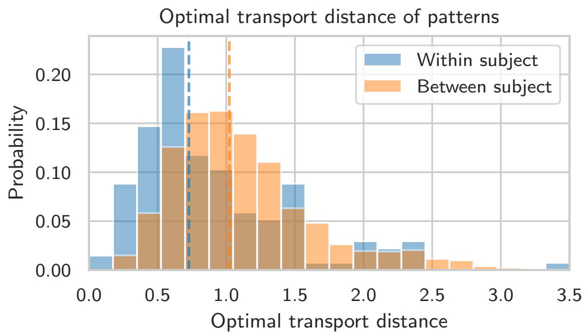

Figure 4: Distribution of the optimal transport distances between spatial patterns within and across subjects. Dashed lines describe median values of 0.728 (within subjects) and 1.023 (between subjects).

the filters. This is due to the fact, that the xDAWN algorithm has almost always found a component representing the P300 evoked potential for each data set (although we had to manually select this component in few cases). As the pattern topology does not differ greatly between subjects and between conditions, we obtain small optimal transport distances. The peak in the within subject histogram values around distances >1.5 are caused by patterns that do not show a typical P300 topology. Overall we conclude, that looking at the pattern alone is not informative enough to describe (dis)similarities in the ERP domain.

To determine the general variability of ERP waveforms and how it affects the Euclidean distance between two ERPs, we calculated the Euclidean distances for three different settings: (1) how ERPs differ between subjects (including different SOA conditions), (2) how ERPs differ within subject under different SOA conditions and (3) how changing the spatial filter but not the underlying data within subjects but across SOA conditions affects the ERP waveform. Hereinafter, we will call this last setting *within subject filter transfer*.

Figure 5 shows the histograms for the three aforementioned settings. Interestingly, the distances resulting from within subject filter transfer are the lowest on average. This indicates that the spatial filters trained on different

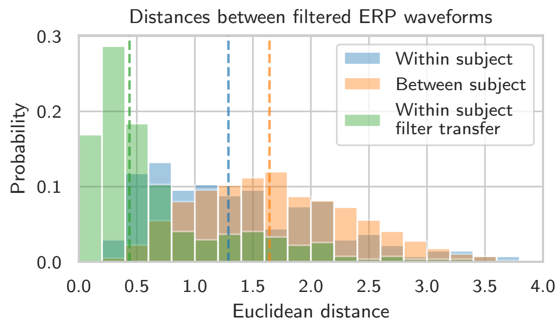

Figure 5: Distribution of the Euclidean distances between ERP waveforms. Blue shows the within subject distances, while orange shows between subject distances. The distances resulting from within subject filter transfer are shown in green. Median values (left to right): 0.437, 1.289, 1.643.

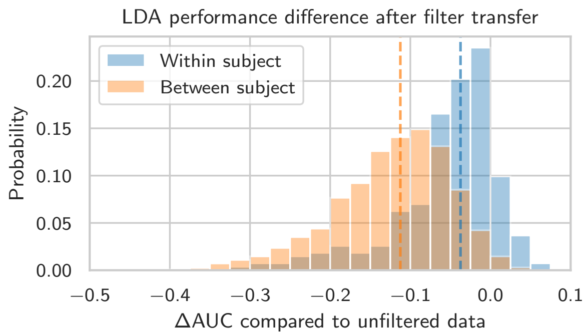

Figure 6: Distribution of ΔAUC for transferring spatial filters within a subject and across subjects. Dashed lines indicate the median values at -0.113, -0.037.

SOAs of one subject can be re-used for other SOA conditions of the same subject without affecting the evoked ERP waveform drastically.

Finally, we quantified how strong the rLDA classifier performance of the xDAWN-filtered data deteriorates when applying a spatial filter trained on a different data set. As shown in Figure 6, the performance losses (provided as ΔAUC values) are much larger when spatial filters from a *different* subject had been used, than when using spatial filters from the same subject. This fits with our previous observations, that spatial filters derived within a subject but across different SOAs still are very similar.

An example for within subject filter transfer is given in Figure 7. We can see that most xDAWN filtered ERP responses look highly similar regardless of which spatial filter we use. Only the filter for SOA 60 seems to extract a different ERP waveform—it generally prolongs the latencies of the ERPs. Applying the filter trained on SOA 60 to the data of SOA 513 even eliminates the P300 ERP completely. This indicates that the filter extracted for this extremely fast SOA value cannot be transferred to the other SOA conditions.

As a contrast, in Figure 8 it is shown how the data recorded for the first four SOA conditions of subject 11 changes when using spatial filters obtained from subject 13. The ERP curves look very different compared

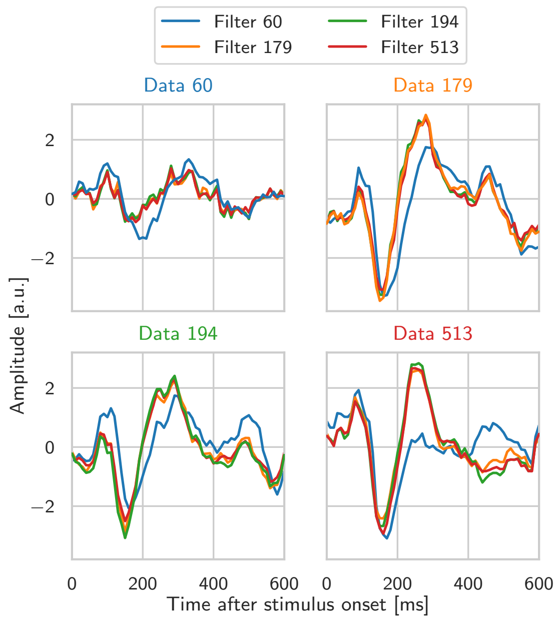

Figure 7: ERP waveforms as a result of filter transfer within subject 11. Each subplot shows a data set on the SOA given in the title. The different colors indicate which spatial filter was applied to the data set. If title and line have the same color, this means that the spatial filter trained specifically on this data set was used to create the filtered ERP signal.

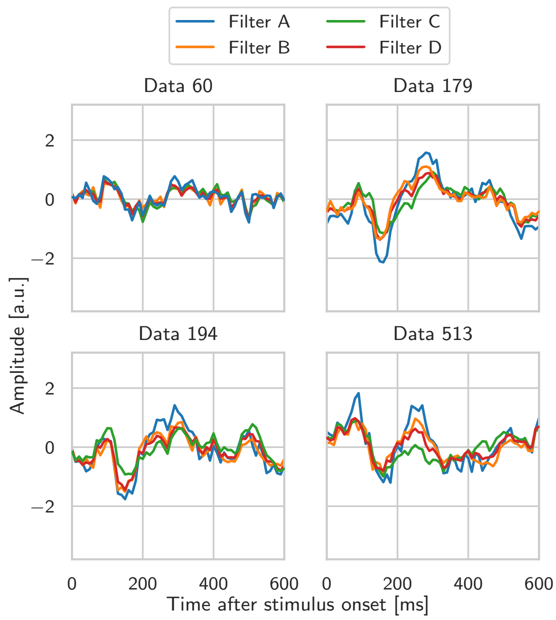

Figure 8: Applied filter transfer from subject 13 to subject 11. Each subplot shows a data set of subject 11 on the SOA given in the title.

to the within subject filter transfer case. Additionally, the difference of ERPs for one data set is much more pronounced by changing the spatial filter.

DISCUSSION AND CONCLUSION

Motivated by the widespread use of data-driven spatial filter methods in the field of BCI, we have described methods to quantify similarities and differences of spatial filters and have applied it to xDAWN filters in the ERP domain. Making use of these methods, we could make a novel contribution to the field by showing that the transfer of spatial filters is feasible for spatial filters trained under different SOA conditions but within one subject—at least if the SOAs are not prohibitively fast. In this context we could report evidence that using solely spatial patterns is not sufficiently informative to characterize spatial components obtained under ERP paradigms.

Our finding may pave the way for a more flexible parameterization of ERP paradigms over multiple sessions of a subject, as it reduces the effort of recording training data for each SOA condition separately in order to get a specialized filter for each condition. An empirical question for future research is, how well this within subject transfer performs if experimental conditions other than SOA are changed, and if the observed successful transfer of filters will be possible also for session-to-session transfer within subjects and between conditions.

In future work, the measures for filter comparison will be examined whether they are applicable also with other spatial filter algorithms and under different experimental paradigms, e.g. CSP, SPoC or ICA in mental imagery tasks or whether there exist measures that can even better reflect transferability properties. Furthermore, an automatic framework that can decide for few data points obtained under varying conditions, whether the resulting spatial filters are transferable or not could help to reduce unnecessary calibration effort.

ACKNOWLEDGMENTS

This work was supported partly by the Cluster of Excellence BrainLinks-BrainTools funded by the German Research Foundation (DFG), grant number EXC 1086, by the DFG project SuitAble (TA 1258/1-1) and by the state of Baden-Württemberg, Germany, through bwHPC and the German Research Foundation (DFG) through grant no INST 39/963-1 FUGG.

REFERENCES

[1] Wolpaw J, Wolpaw E. Brain-computer interfaces: Principles and practice. Oxford University Press, USA (2012).

[2] Farwell LA, Donchin E. Talking off the top of your head: Toward a mental prosthesis utilizing event-related brain potentials. Electroencephalography and Clinical Neurophysiology. 1988;70(6):510–523.

[3] Schreuder M, Rost T, Tangermann M. Listen, you are writing! Speeding up online spelling with a dynamic auditory BCI. Frontiers in Neuroscience. 2011;5.

[4] Ang KK, Guan C. Brain-computer interface in stroke rehabilitation. 2013.

[5] Blankertz B, Tangermann M, Vidaurre C, *et al.* The Berlin brain–computer interface: Non-medical uses of BCI technology. Frontiers in neuroscience. 2010;4:198.

[6] Krauledat M, Grzeska K, Sagebaum M, *et al.* Playing pinball with non-invasive BCI. In: Advances in neural information processing systems. 2009, 1641–1648.

[7] Nitschke JB, Miller GA, Cook EW. Digital filtering in EEG/ERP analysis: Some technical and empirical comparisons. Behavior Research Methods, Instruments, & Computers. 1998;30(1):54–67.

[8] Blankertz B, Kawanabe M, Tomioka R, Hohlefeld F, Müller K-R, Nikulin VV. Invariant common spatial patterns: Alleviating nonstationarities in brain-computer interfacing. In: Advances in neural information processing systems. 2008, 113–120.

[9] Rivet B, Souloumiac A, Attina V, Gibert G. xDAWN algorithm to enhance evoked potentials: Application to brain–computer interface. IEEE Transactions on Biomedical Engineering. 2009;56(8):2035–2043.

[10] Winkler I, Brandl S, Horn F, Waldburger E, Allefeld C, Tangermann M. Robust artifactual independent component classification for BCI practitioners. Journal of neural engineering. 2014;11(3):035013.

[11] McFarland DJ, McCane LM, David SV, Wolpaw JR. Spatial filter selection for EEG-based communication. Electroencephalography and Clinical Neurophysiology. 1997;103(3):386–394.

[12] Dähne S, Meinecke FC, Haufe S, *et al.* SPoC: A novel framework for relating the amplitude of neuronal oscillations to behaviorally relevant parameters. NeuroImage. 2014;86:111–122.

[13] Polich J. Task difficulty, probability, and interstimulus interval as determinants of P300 from auditory stimuli. Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section. 1987;68(4):311–320.

[14] Kindermans P-J, Tangermann M, Müller K-R, Schrauwen B. Integrating dynamic stopping, transfer learning and language models in an adaptive zero-training ERP speller. Journal of neural engineering. 2014;11(3):035005.

[15] Haufe S, Meinecke F, Görgen K, *et al.* On the interpretation of weight vectors of linear models in multivariate neuroimaging. NeuroImage. 2014;87(0):96 –110.

[16] Krauledat M, Tangermann M, Blankertz B, Müller K-R. Towards zero training for brain-computer interfacing. PloS one. 2008;3(8):e2967.

[17] Xu Y, Wei Q, Zhang H, *et al.* Transfer learning based on regularized common spatial patterns using cosine similarities of spatial filters for motor-imagery BCI. Journal of Circuits, Systems and Computers. 2018:1950123.

[18] Rubner Y, Tomasi C, Guibas LJ. The earth mover's distance as a metric for image retrieval. International journal of computer vision. 2000;40(2):99–121.