Using within-subject pattern classification to understand lifespan age differences in oscillatory mechanisms of working memory selection and maintenance

Julian D. Karch, Myriam C. Sander, Timo von Oertzen, Andreas M. Brandmaier, Markus Werkle-Bergner

PII: S1053-8119(15)00329-8
DOI: doi: 10.1016/j.neuroimage.2015.04.038
Reference: YNIMG 12164

To appear in: *NeuroImage*

Received date: 2 October 2014
Accepted date: 19 April 2015

Please cite this article as: Karch, Julian D., Sander, Myriam C., von Oertzen, Timo, Brandmaier, Andreas M., Werkle-Bergner, Markus, Using within-subject pattern classification to understand lifespan age differences in oscillatory mechanisms of working memory selection and maintenance, *NeuroImage* (2015), doi: 10.1016/j.neuroimage.2015.04.038

Using within-subject pattern classification to understand lifespan age differences in oscillatory mechanisms of working memory selection and maintenance

Julian D. Karch[a,*], Myriam C. Sander[a], Timo von Oertzen[a,b], Andreas M. Brandmaier[a,1], Markus Werkle-Bergner[a,1,*]

[a] *Center for Lifespan Psychology, Max Planck Institute for Human Development, Berlin, Germany*
[b] *Department of Psychology, University of Virginia, Charlottesville, USA*

Abstract

In lifespan studies, large within-group heterogeneity with regard to behavioral and neuronal data is observed. This casts doubt on the validity of group-statistics-based approaches to understand age-related changes on cognitive and neural levels. Recent progress in brain–computer interface research demonstrates the potential of machine learning techniques to derive reliable person-specific models, representing brain behavior mappings. The present study now proposes a supervised learning approach to derive person-specific models for the identification and quantification of interindividual differences in oscillatory EEG responses related to working memory selection and maintenance mechanisms in a heterogeneous lifespan sample. EEG data were used to discriminate different levels of working memory load and the focus of visual attention. We demonstrate that our approach leads to person-specific models with better discrimination performance compared to classical person-nonspecific models. We show how these models can be interpreted both on an individual as well as on a group level. One of the key findings is that, with regard to the time dimension, the between-person variance of the obtained person-specific models is smaller in older than in younger adults. This is contrary to what we expected because of increased behavioral and neuronal heterogeneity in older adults.

Keywords: prediction, lifespan age differences, EEG, single-trial analysis

*Corresponding author
Email addresses: karch@mpib-berlin.mpg.de (Julian D. Karch), werkle@mpib-berlin.mpg.de (Markus Werkle-Bergner)

[1] Equal contribution

ANOVA analysis of variance

BAC balanced accuracy

1. Introduction

Across the lifespan, working memory (WM) performance and the underlying mechanisms of selection and maintenance undergo a tremendous change (Sander et al., 2012a). On the behavioral level, an increase in performance across childhood with a peak in young adulthood is followed by decline with advancing age. On the neural level, WM performance depends, among others, on rhythmic neural activity in the alpha band (8–12 Hz) (e.g., Freunberger et al., 2011). In particular, posterior alpha (8–12 Hz) power changes have been related to attentional shifts and WM load during retention (e.g., Obleser et al., 2012; Sander et al., 2012b; Sauseng et al., 2009). Given low WM performance in children and older adults, altered mechanisms of rhythmic neural processing are to be expected in these age groups. However, evidence on age-related changes in rhythmic neural activity related to WM maintenance and selection is scarce, and results are mixed. Some studies have found evidence for age-differential attentional effects on the modulation of alpha power (Sander et al., 2012b), whereas others did not observe any attention-related modulation in older adults (Vaden et al., 2012).

One possible explanation for the mixed results may be increased heterogeneity in children and older adults. Implicit to inferring cognitive and neural processes from group data is the strong and necessary (but not sufficient) assumption that the process under investigation is equivalent for every group member. Hence, group homogeneity is crucial for making links between cognitive and neural data. Consequently, conclusions about processes from the sample-specific level to the person-specific level are valid only under the ergodic assumption (Molenaar, 2004). Ergodicity assumptions are typically violated for developmental processes, and, in these cases, it is necessary to base analyses on intraindividual variability rather than on interindividual change. As a result, inferences on the individual level may be diluted if not meaningless (e.g., Hayes, 1953; Molenaar and Campbell, 2009; Nesselroade et al., 2007; Siegler, 1987; Voelkle et al., 2014).

Thus, increased interindividual variability in developmental populations may cast doubt on the validity of group statistics, and calls for the development of analyses based on intraindividual instead of interindividual

---

BCI brain–computer interface

CI confidence interval

CSP common spatial pattern

EEG electroencephalography

ICA independent component analysis

LDA Linear Discriminant Analysis

LLDA Ledoit's Linear Discriminant Analysis

PCA principal component analysis

WM working memory

models (e.g., Nesselroade et al., 2007). The high reliability and reproducibility of behavioral, as well as structural and functional brain measures in younger adults suggests reasonable homogeneity of this group. However, widespread changes in brain chemistry, anatomy, and functionality are documented, especially for child development and aging (e.g., Bäckman et al., 2006; Raz et al., 2005). These changes are typically accompanied by increased heterogeneity of functioning in behavioral tasks (e.g., Astle and Scerif, 2011; Nagel et al., 2009; Werkle-Bergner et al., 2012). Hence, especially for cross-sectional comparative studies, one may question whether meaningful between-group comparisons are feasible knowing that the within-group heterogeneity is not equal across lifespan samples.

Here, we suggest using within-subject pattern classification to better understand lifespan age differences in oscillatory mechanisms of WM selection and maintenance. More precisely, we propose a two-step approach: First, person-specific models that explore the expected relationship between brain responses and experimental conditions on the individual level are estimated. For example, in the context of multivariate time-series analyses, this can be achieved by estimating person-specific parameter estimates for a given model of interest (Nesselroade et al., 2007). In a second step, invariance of the parameter estimates can be tested within and across groups (Boker et al., 2009). Inspired by those considerations, we propose a formal approach to derive person-specific models for the identification of differential brain–behavior links in lifespan samples.

In multi-channel electroencephalographic (EEG) recordings of brain signals over time, a person-specific model that carries maximal information about the discrimination of experimental conditions can vary across several parameters: first and foremost, across time, duration, channel, and frequency. The model space spanned by these parameters is necessarily large and finding the most informative model is far from trivial. Therefore, we formalize the problem as a classification task and employ multivariate pattern classification algorithms (also know as multivariate pattern analysis (MVPA) in the (f)MRI literature (Norman et al., 2006)) in combination with a precisely tailored preprocessing chain to obtain a solution. Similar approaches have been dominating brain–computer interface (BCI) research. However, in many of these applications, the predictive accuracy is the primary target of the procedure rather than the inference about the underlying processes. Our framework can be regarded as a supporting tool in the recursive interplay of theory-guided and exploratory analysis of neuroimaging data that assists researchers in hypothesis generation and theory building by extracting stable patterns from data (cf. Brandmaier et al., 2013). In our work, we place a particular emphasis on interindividual differences as typically encountered in lifespan and aging research.

To test the applicability of our approach, we re-analyze data from a lifespan study that targeted brain oscillatory mechanisms for WM selection and WM maintenance in a lifespan sample including children, younger, and older adults (Sander et al., 2012b). The study used a color change–detection task (Vogel and Machizawa, 2004), in which participants were cued to attend to either the left or the right hemifield and asked to remember the colors of varying numbers of items. Hence, by design, it is possible to identify modulations of rhythmic neural responses that (a) relate to the attentional focus and (b) reflect the varying levels of WM load. We operationalized (a) attentional focus as the hemifield to which spatial attention should be shifted and (b) WM load as the amount of items to be remembered in a change–detection task. Hence, in a first step, we set out to predict the focus of visuospatial attention based on changes in (posterior) alpha power. We

will refer to this as attentional focus prediction in the following. Given the robust relation of posterior alpha power modulations and attention shifts (e.g., Kelly et al., 2006; Sander et al., 2012b; Worden et al., 2000), this analysis was intended as a validation step of our classification approach (e.g., Bahramisharif et al., 2010; van Gerven and Jensen, 2009; Kelly et al., 2005, for previous BCI approaches). This part of the study aimed to demonstrate the feasibility of deriving person-specific models with varying spatio-temporal information in groups of children, younger, and older adults. In a second step, we aimed to predict information maintained in WM based on single-trial modulations of neural activity in the alpha range. We will refer to this as WM load prediction in the following. Previous studies have successfully demonstrated load modulations of lateralized alpha power activity (Sauseng et al., 2009). However, given that studies demonstrating the possibility of WM load prediction from scalp EEG recordings are scarce (but see Roux et al., 2012, for an analysis of source activity in pre-identified regions), this part represents an extension of the applicability of our classification approach.

2. Material and methods

2.1. Identifying person-specific models: The classification approach

The core idea of our framework is the derivation of person-specific models that optimally discriminate between behavioral conditions and, thus, allow evaluation of the neural underpinnings of interindividual differences in behavioral responses. Here, the term *model* is used to refer to both a class of models (representing a particular functional form with unknown parameters) and a particular instance of a model with parameters estimated from data. Whenever the distinction is not clear from the context, we refer to the latter as an *estimated model*. Person-specific models are estimated models selected from a set of candidate models that vary across multiple dimensions of the observed data space. In electroencephalography (EEG), this space typically entails electrode channels, time points, and/or frequencies; but our considerations generally apply to any spatio-temporal method of brain imaging. Candidate models can be derived from a template model class and vary parametrically according to multiple dimensions, first and foremost, to the spatio-temporal segments of the original data they are exposed to. In particular, models operate on different time windows and on subsets of channels or their geometric projections. In the remainder of this subsection we will describe the proposed framework to estimate person-specific models.

In the following, the number of measured variables per sample will be denoted by M and the number of samples per individual will be denoted by T, as those typically refer to samples ordered in time. For each person, a data set $(\mathbf{x}_t, y_t) \in D$ with $t \in \{1, \ldots, T\}$ is measured, which is a set of tuples of observed brain responses $\mathbf{x}_t \in \mathbb{R}^M$ and a corresponding dichotomous target variable $y_t \in \{0, 1\}$ that typically corresponds to a given external condition, task, or state. A candidate model, mapping brain responses to the target variables, can then be conceived as a θ-parameterized function $f_\theta(\mathbf{x}) = y$, linking the observed neural responses $\mathbf{x}_t$ and behavioral states y_t. The specific parameters θ can be estimated by minimizing a loss function on data (usually called the training set). Each estimated model can then be evaluated with respect to its accuracy in predicting a behavioral condition from brain responses, whereby selection of the best model is carried out

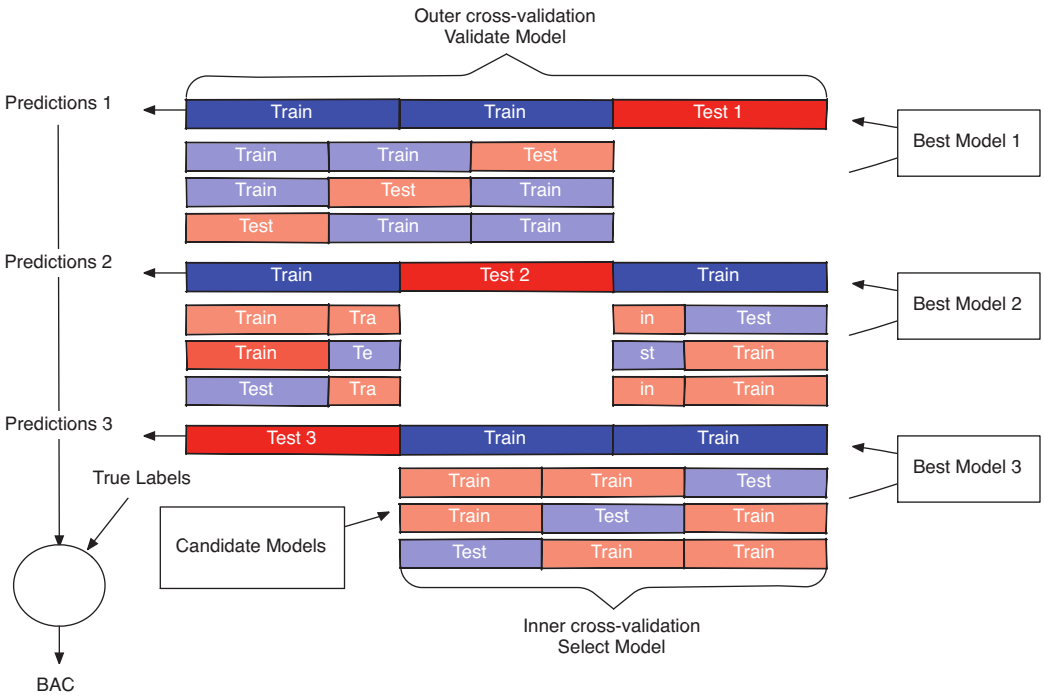


Figure 1: **Nested cross-validation.** Schematic representation of the nested cross-validation procedure for model selection and evaluation. The schema shows a three-fold nested cross-validation procedure instead of the 10-fold procedure we employed. The model selection takes place in the inner loop (light blue and light red), while model evaluation of the best inner model is performed in the outer loop (dark blue and dark red). The evaluation metric is the balanced accuracy (BAC).

for each person separately. We propose to use the balanced accuracy (BAC), a loss function accounting for unbalanced target variables that are often encountered in EEG data sets, as the performance measure for each candidate model. The BAC is the average of the accuracies obtained for each target variable state (condition) (Brodersen et al., 2010). This metric allows us to select the best of all competing models and interpret the idiosyncratic brain–space information of that model as person-specific information.

To avoid an overoptimistic bias by confounding parameter estimation and model evaluation (Kriegeskorte et al., 2009; Stone, 1974), we estimate the BAC for a given candidate model using 10-fold stratified cross-validation (Kohavi et al., 1995). The resulting person-specific models can be interpreted as both a measure of interindividual heterogeneity in the sample and as a parsimonious indicator of the location and magnitude of these interindividual differences in brain space. In the remainder of this subsection we will describe the procedure of choosing the best person-specific model in more detail (see also Fig.1).

Selection of an optimal model for each person from a set of candidate models entails a cherry picking problem. This cherry picking causes an overoptimistic accuracy estimate if the best accuracy of the model selection phase is reused as the estimate for the overall inner performance of the selected model (Stone, 1974; Varma and Simon, 2006). To obtain an unbiased BAC estimate, we separately use 10-fold stratified cross-validation for model selection and model evaluation, which leads to a nested cross-validation procedure with an inner and an outer loop. Nested cross-validation is the de-facto standard procedure for performance evaluation in BCI research (Lemm et al., 2011). The cross-validated model *selection* takes place in the inner loop, while cross-validated model *evaluation* is performed in the outer loop. To this end, we randomly partition a data set into ten exhaustive and mutually exclusive subsets. For each of the ten outer cross-validation runs, a temporary training set is formed by leaving out one of the ten folds whereas a temporary test set is formed by the remaining folds. The temporary outer training set forms the basis for the inner cross-validation runs aiming to determine the temporary best model (see below). The performance of the latter is evaluated on the current outer test set. To identify the best model in each run of the outer loop, an inner cross-validation procedure is used. Here again, the available data (i.e., the current outer training set) are split into ten exhaustive and mutually exclusive subsets (i.e, the inner training and test sets for model selection). In each run of the inner loop, the parameters for the candidate models are estimated on the temporary inner training set and their performance is evaluated on the current inner test set. The overall performance estimate for a given candidate model is obtained by aggregating the test set performance across all 10 iterations of the inner loop. The candidate model with the highest average BAC is selected as the current best model and its performance is evaluated on the remaining outer test set.

Technically, the outer cross-validation does not estimate the BAC of the best candidate model but rather the BAC of the model selection procedure as a whole. Particularly, in each fold of the outer cross-validation, a different model may be selected as the best one. In other words, nested cross-validation provides an unbiased estimate of the expected BAC when applying the model selection of the inner cross-validation to the whole data set of one person. Put differently, we estimate the expected accuracy of picking a model from our candidate set based on cross-validated predicted accuracy. Consequently, after the nested cross-validation evaluation, the model selection phase (i.e., only the logic of the inner loop) is applied to the complete sample

to obtain the best person-specific model. As a final step, the parameters θ of the best model are estimated using the complete data set. The BAC estimate obtained by nested cross-validation is a conservative estimate for the performance of the model using parameters derived from the whole data set.

2.2. Candidate models

In the following, we describe the set of candidate models from which the model selection framework determines the best model for each person. Foremost, our goal is to predict the attentional focus as well as WM load based on differences in the power of neural activity in the alpha range (e.g., Jensen et al., 2002; Sander et al., 2012b; Sauseng et al., 2009). A further challenge is the identification of the most discriminative time window per person. To achieve the latter, only data from within a single time window are considered for each candidate model. Hence, the models differ with regard to the onset and the duration of the employed time window.

To capture differences in signal power we chose the CSP method (Müller-Gerking et al., 1999; Ramoser et al., 2000). CSP finds a transformation matrix $\mathbf{C}$, mapping the EEG channels onto a set of component projections such that the variances of the resulting time series discriminate optimally between conditions (Ramoser et al., 2000). In order to find the respective transformation matrix $\mathbf{C}$, the eigenvalue decomposition of the mean between-channel covariance matrix is computed. Thus, CSP requires an invertible mean between-channel covariance matrix. If the goal is to discriminate conditions 0 and 1, the projected CSP components are ordered such that the variance of the first component is maximal for condition 0 while being minimal for 1. Vice versa, the last projected component has maximal variance for condition 1 while it is minimal for 0. Hence, the respective components from both ends form complementary pairs with regard to condition prediction. Like other dimensionality reduction approaches such as principal component analysis (PCA), a subset of components can be selected. Due to their complementarity, CSP components are typically picked in pairs, with each subsequent pair adding less predictive information to the overall task. In the following, we refer to the number of these CSP filter pairs as k. To reiterate, a classifier relying on a single filter pair ($k = 1$) bases its prediction on two projected components, for each of which the variance is most informative about the respective target outcome. Consequently, the variances of the EEG projected onto the CSP filters are used as features, that is, as classifier input.

Classification is done by Linear Discriminant Analysis (LDA). LDA is a linear classifier, meaning that, its decision function is of the form $\text{sign}(\mathbf{w}^\top \mathbf{x} + c)$, where $\mathbf{x}$ are the features, $\mathbf{w}$ is called weight vector, and c bias. For the training of LDA, the feature means and covariance matrices for each class have to be estimated. If the number of samples is large in comparison to the number of features, the sample covariance matrix is a sufficiently precise estimate of the population covariance matrix. For the typical EEG classification problem, the number of samples is roughly equal to or smaller, than the number of features. Hence, the sample covariance matrix is systematically biased (Friedman, 1989). To correct for this bias, regularization is commonly used (Friedman, 1989). The regularization hyper-parameter was set by the analytical solution of Ledoit and Wolf (2004). We refer to the resulting classifier as Ledoit's Linear Discriminant Analysis (LLDA). The combination of CSP and LDA is commonly used to realize BCIs based on rhythmic neural activity (e.g.

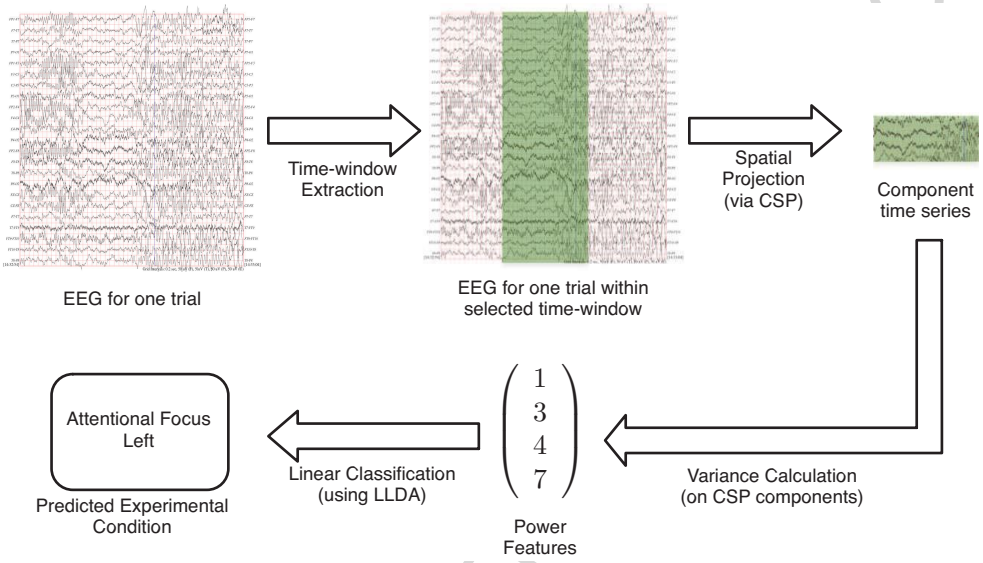


Figure 2: **Decision function.** Graphical description of the decision function learnt by the proposed candidate models. Prior to applying this decision function, the EEG is preprocessed. First, the EEG trial is reduced by only considering the signals within a selected time window. After that, the $2k$ selected components of the projection matrix $\mathbf{C}$, as learnt by CSP, are applied to the remaining signals. This yields $2k$ component time series. Next, the variance of each of these time series is calculated. The resulting power features are classified by the linear decision function as learnt by Ledoit's Linear Discriminant Analysis (LLDA).

Blankertz et al., 2010; Fazli et al., 2009).

The different candidate models were derived from the following settings: Duration of the time windows: $\{100, 150, 200, \ldots, 700\}$ ms, onset of the time window: $\{0, 33, 66, 99, \ldots, L\}$ ms relative to the onset of the memory array (see Section 3 for a detailed description of the trial design). The limit L for the onset depended on the duration of the time window. It was chosen such that the latest time windows did not contain signals later than 1000 ms after the onset of the memory array. This was done in order to prevent inclusion of EEG activity evoked by the onset of the test stimulus. As candidates for the number k of the CSP filter pairs we employed $\{2, 3, 4, 5, 6\}$. These candidates were motivated by the rationale of exploring the values around the recommendation to use $k = 3$ by Blankertz et al. (2008). The full grid of all possible combinations of settings was explored, leading to 5 (filter pairs) $\times$ 244 (time windows) = 1220 different candidate models.

2.3. Spatial interpretation of the best estimated model

In oder to interpret the spatial information we obtained filters and patterns (Bießmann et al., 2012; Parra et al., 2005) corresponding to our estimated predictive model. Filters and patterns assume that the observed data obey a linear measurement model. That is, we assume a set of sources mapped to the observed values by a linear transformation. For raw EEG potentials it is generally assumed that a linear measurement model holds true. A pattern describes the contribution of one source to all electrodes (forward model). A filter describes the linear reconstruction of one source given the observed data (backward model). For a linear classifier the filter simply corresponds to the weight vector $\mathbf{w}$. For more details about filters and patterns, and their relationship, see AppendixA.1. In contrast to previous work that employed classifiers to obtain person-specific filter and patterns (Parra et al., 2005; Philiastides and Sajda, 2006), we did not use raw EEG potentials as features for our classifier. Instead, the model class that we proposed in Subsection 2.2 yields a linear classifier with the variances of the component time series as features. This is equivalent to a linear classifier with all entries of the covariance matrix between all channels as features (see AppendixA.2), and thus produces a filter with $\frac{(M+1)M}{2}$ entries, where M is the number of channels.

It follows directly from the linear model assumption for EEG data that the observed covariance data does not comply with a linear model (see AppendixA.4). The resulting filter still corresponds to the classifier weights $\mathbf{w}$ and the patterns can also still be computed (see AppendixA.3). But their interpretation as forward and backward models is no longer valid. However, there is still a meaningful interpretation for both filter and pattern. The weight w_p of the filter corresponding to the p-th feature expresses that an increase of the p-th feature by one increases the classification score $\mathbf{w}^\top \mathbf{x} + c$ by w_p. Thus, for positive/negative weights, higher values increase/decrease the classification score and hence, the support for the positive/negative class. Positive class refers to the class that is predicted if $\text{sign}(\mathbf{w}^\top \mathbf{x} + c) \geq 0$. For our prediction tasks "Attention left" and "Low WM load" are the positive classes and "Attention right" and "High WM load" are the negative classes (see Subsection 2.4.2).

For the pattern the opposite logic applies. If the classification score increases by one, the expected observed value of the p-th feature changes by a_p. As pattern and filter are not invariant to the scaling of the features we normalized both. For illustration purposes we show only the weights corresponding to the variances.

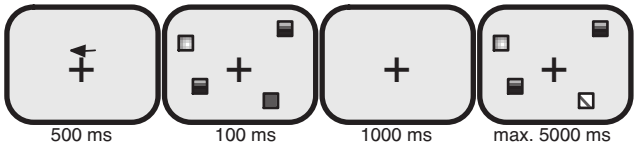


Figure 3: **Experimental paradigm.** Each trials starts with the presentation of a cue, indicating the relevant hemifield. The memory set is presented for 100 ms and followed by a fixed retention interval of 1000 ms. The probe display is shown until a response is given. Different patterns of the squares represent different colors. Adapted with permission from Sander et al. (2012b)

Note, however, that for classification and for calculation of the pattern we included all terms, omitting the covariances proved considerably worse (see Section 3.4).

2.4. Study design

We used the approach proposed above for the analysis of two data sets from a study that investigated lifespan age differences in WM in groups of children, younger adults, and older adults (Sander et al., 2012b). Details on procedures, task design, and EEG data recording can be found there. For convenience, we outline the study design as it is important for the understanding of our analysis approach as well as the results.

2.4.1. Participants

For the present study, we used data from 22 children ($M_{\text{age}} = 11.9$ years, range 10–13 years), 12 younger adults ($M_{age} = 24.2$ years, range 20–25 years), and 22 older adults ($M_{age} = 73.3$ years, range 70–75 years). Given that groups of children and older adults are typically more heterogeneous in comparison to younger adults, sample sizes for these two age groups were larger in the initial report. The initial sample included 31 children, 19 younger adults, and 31 older adults. Details about exclusion criteria and descriptive marker tests documenting the age typicality of the sample can be found in Sander et al. (2012b). The ethics committee of the Max Planck Institute for Human Development, Berlin, approved the study.

2.4.2. Experimental paradigm

The exact procedures are described in Sander et al. (2012b). In short, a hemifield version of the change–detection task (Vogel and Machizawa, 2004) was used to probe age differences in visual WM capacity (see Fig. 3). Memory arrays of colored squares were presented to the participants for 100 ms or 500 ms. Targets were defined as the squares presented in the hemifield indicated by a centrally placed cue before each trial. To keep the difficulty of the task comparable for the different age groups, memory arrays of 2, 4, or 5 targets were presented to younger adults while they involved 2, 3, or 4 targets for the older adults and children. Besides the number of targets, the experimental procedure was identical for all age groups. For the present report, we only analyzed the common set sizes 2 and 4. In the following, they will be referred to as low (2) and high (4) levels of WM load. Trials were presented in four blocks. The first two and the last two blocks were always presented at the same presentation time, and the order of the presentation time was counterbalanced across

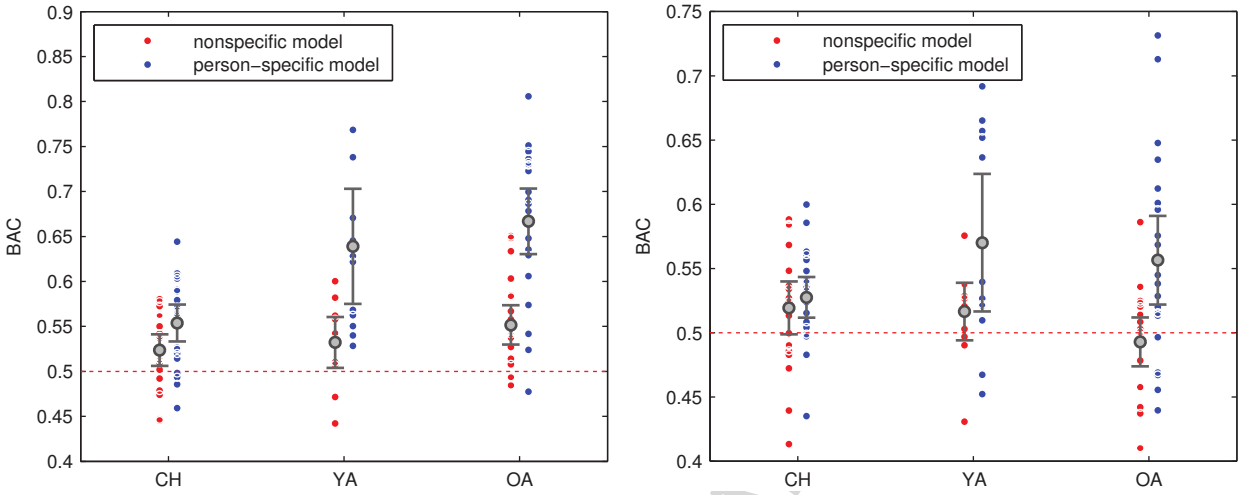
participants. Set sizes were randomized within blocks. After a retention interval of 1000 ms, a probe array of colored squares was shown and participants had to indicate whether all the colors of the probe array's targets were identical to the memory array or whether one of the squares had changed in color. We only considered subsegments of the time segment starting with the presentation of the memory array and ending 1000 ms after that for our analysis (see Subsection 2.2). The presentation of each block started with 12 practice trials. Then, each participant completed 360 trials of varying set size. Set sizes were randomized within blocks. After each block, participants got feedback about the accuracy of their responses. Given that we assumed older adults and children would have additional difficulties with a cued hemifield presentation, we always presented the cue for 500 ms and showed it until the memory array was presented to minimize cue-related memory load. We also blocked the cue direction for 30 consecutive trials to prevent a task-switching situation that could differentially affect the age groups (e.g., Kray and Lindenberger, 2000).

2.5. Preprocessing

For preprocessing, the EEG was re-referenced to the two mastoid channels. Afterwards, an independent component analysis (ICA) was used to correct for remaining eye blink, noise, and muscle activity (Jung et al., 2000). Independent components representing artifactual sources were visually identified and removed from the data. Thereafter, trials with an incorrect response were removed. As the last preprocessing step PCA was used to project the EEG onto the principal components that explained 99% of the variance. This was done to restore the invertibility of the mean between-channel covariance matrix, which was violated by the removal of independent components in the previous preprocessing step. As explained above, the CSP method that we chose as part of our candidate models (see Subsection 2.2) requires an invertible mean between-channel covariance matrix. An almost identical but more elegant way to restore invertibility would have been to simply project the EEG onto the retained independent components. We chose the ICA, PCA combination for practical reasons. As the exact location of individual frequency bands may change across the lifespan (Klimesch, 1999), the individual alpha peak frequency was estimated for each individual participant based on independently assessed resting state data. To determine the individual alpha peak frequency we computed power spectra for eyes-closed resting state data and averaged them across all occipito-parietal electrodes. The individual alpha peak frequency was then defined as the maximum peak of the averaged power spectra between 7 and 13 Hz (see Sander et al., 2012b, for more details). The cut-off frequencies for band-pass filtering into the alpha frequency ranges were determined in relation to individual alpha frequency based on suggestions by Doppelmayr et al. (1998).

2.5.1. Data analysis

Previous studies have examined load modulations of lateralized alpha power activity at 100 ms presentation times (Sauseng et al., 2009). Therefore, the analyses presented in this study are focused on this presentation time condition, which is the standard condition used in change detection paradigms (e.g., Luck and Vogel, 1997; Vogel and Machizawa, 2004). Analyses were conducted with custom-made MATLAB code (The MathWorks Inc., Natick, MA, USA) based on the Fieldtrip software package (Oostenveld et al., 2010).



(a) Attentional focus (b) WM load

Figure 4: **Performance evaluation.** Mean BAC and CIs for the person-specific (blue dots) and the nonspecific model (red dots). Shown for the attentional focus (a) and WM load (b) classification for each of the three age groups separately. Each dot describes the BAC of one person. The box plots denote the respective mean values and 95% CIs. The red dotted horizontal line describes the BAC as expected by the null hypothesis. "CH" stands for children, "YA" for younger adults, and "OA" for older adults.

3. Results

First, we validate our model class by comparisons against chance and a theory-driven nonspecific model class. After that, we present detailed person-specific results for selected participants, followed by group results. Finally, the proposed model class is compared against simpler ones.

3.1. Performance evaluation against chance and the best nonspecific model

To validate our framework and the set of candidate models, we will demonstrate that our approach resulted in a classification performance significantly different from chance in each age group. Moreover, we will show that the accuracy is higher than for a conventional theory-driven model.

To test our approach against chance, we compared it against a model that guesses class membership on each trial. Such a decision function guessing for each trial will, in the limit, achieve a BAC of 0.5. Therefore, we concluded that the prediction accuracy of the person-specific models within an age group was reliably different from chance if the respective 95% CI of the mean BAC did not include 0.5. We calculated CIs based on the t-distribution, for each age group. To test for univariate normality of BACs within each group the Shapiro-Wilk test (Royston, 1995) and quantile-quantile plots were used. For all CIs the p-values for the Shapiro-Wilk test were larger than 0.05. Visual inspection of the quantile–quantile plots also suggested that

the data were normally distributed. Furthermore, ceiling effects, typical for accuracy data, were not present. We therefore believe that the CIs based on the t-distribution were a reasonable choice for quantifying the reliability of our estimates. We conclude that the prediction accuracy of the person-specific models within an age group is reliably different from chance if the p-value of the one-sided t-test lies below 0.05. As Fig. 4 shows for all age group–task combinations the CIs of the person-specific models did not include 0.5. Consequently, the p-value of the one-sided t-test was below 0.05 for all age groups. Hence, the person-specific models allowed for a reliable classification of attentional foci and WM load based on neural activity in the alpha frequency range. As effect size we report the difference between the mean BAC and 0.5. The effect sizes for children, younger adults, older adults respectively were: attentional focus prediction: 0.0538, 0.1390, 0.1669; and WM load prediction: 0.0276, 0.0701, 0.0565.

To compare our person-specific approach against conventional theory-driven, nonspecific analysis schemes that assume the same model for each person, we needed an appropriate a-priori model. Previous studies have observed effects of shifted attention in early time windows (e.g., Freunberger et al., 2008), whereas effects of load were usually reported from the maintenance period of the change–detection task (Grimault et al., 2009; Vogel and Machizawa, 2004). Thus, we used 0–400 ms as the theory-driven time window for attentional focus prediction and 400–1000 ms for WM load prediction. In line with conventional analysis schemes (e.g., Pfurtscheller and Aranibar, 1977), the variances of the bandpass-filtered signals were used as features for the prediction instead of CSP. Conventional analysis techniques are univariate in nature. Therefore, the equivalent of a weight vector (or any other function that integrates evidence), as needed for linear prediction, is not available. Therefore, we used LLDA to learn the weight vector and classify the power features. We again used 10-fold cross-validation to evaluate its performance. To obtain the single best weight vector we applied cross-validation to the data of all participants simultaneously, but separately for both tasks. That is, we treated the data of all persons as a single person. As a result, we obtained an unbiased BAC estimate for the best estimated nonspecific model, explicitly implementing the expectation that the same neural mechanisms are present in all participants (Danziger, 1990; Molenaar and Campbell, 2009; Nesselroade et al., 2007).

First, to examine whether there is an effect across all age groups of using the nonspecific or the person-specific model, we performed a repeated measures analysis of variance (ANOVA) with age group as the between factor and model type (person-specific or nonspecific) as the within factor. For the attentional focus prediction the main effect of model ($F(1, 53) = 61.44, p = 2 \times 10^{-10}$) as well as the interaction effect of model and age group were significant ($F(2, 53) = 7.62, p = 0.0012$). The interaction was ordinal, thus just revealing differences in the strength of the main effect between age groups. The main effect of model was positive. Hence, across all age groups, the person-specific model improved the BAC for the attentional focus prediction. For the WM load prediction the main effect of model ($F(1, 53) = 14.80, p = 0.0003$) was significant. The interaction effect of model and age missed the conventional significance level ($F(2, 53) = 3.01, p = 0.0577$). Hence, the positive main effect of model is interpretable. Thus, across all age groups, the person-specific model improved the BAC for the WM load prediction.

Fig. 4 depicts the mean BAC for the nonspecific as well as the person-specific model. With regard to the classification of the attentional focus, for all three age groups the person-specific model was more accurate

	Nonspecific	Specific	p-value	Mean Difference
Children	$[0.5061, 0.5412]$	$[0.5332, 0.5744]$	0.0258	0.0302
Younger adults	$[0.5040, 0.5605]$	$[0.5749, 0.7031]$	0.0018	0.1068
Older adults	$[0.5298, 0.5735]$	$[0.6305, 0.7033]$	9×10^{-8}	0.1152

(a) Attentional focus

	Nonspecific	Specific	p-value	Mean Difference
Children	$[0.4988, 0.5400]$	$[0.5117, 0.5435]$	0.2520	0.0082
Younger adults	$[0.4942, 0.5390]$	$[0.5166, 0.6236]$	0.0313	0.0535
Older adults	$[0.4739, 0.5119]$	$[0.5220, 0.5910]$	0.0015	0.0636

(b) WM load

Table 1: Mean BAC and CIs for the person-specific and the nonspecific model. Corresponding p-values from the paired one-sided t-test. These CIs are visualized in Fig. 4

than the nonspecific model. The respective 95% CIs, effect sizes (mean difference), and p-values obtained by the paired one-sided t-test are shown in Table 1a. Similar results were obtained for the WM load prediction. Here, with exception of the children, the person-specific model was more accurate than the nonspecific model. The respective 95% CIs, effect sizes (mean difference), and p-values obtained by the paired one-sided t-test are shown in Table 1b.

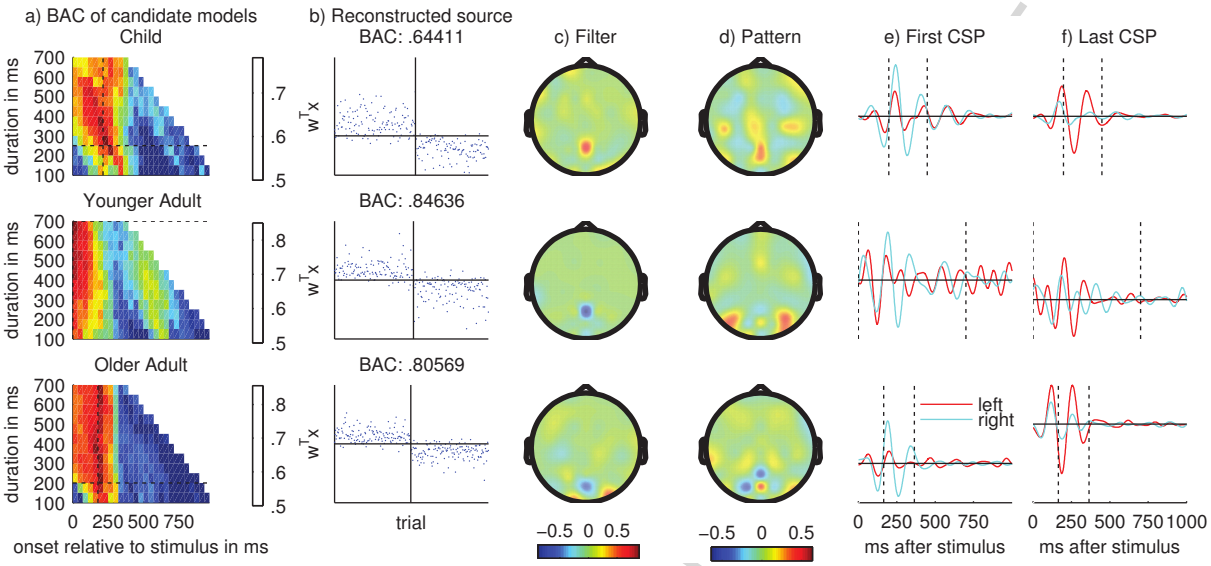
Cross-validation results are not necessarily well approximated by a normal distribution, and the variances for two classification methods are different depending on the means of the two distributions. Hence, a t-test is only an approximation of a valid test – in both cases, the test against chance and the comparison of the models against each other. However, in our case the large effect sizes justify the t-test. To substantiate the test further, we repeated the test with a permutation test (Nettleton and Doerge, 2000) for the case with the smallest effect size (WM load prediction in children with person-specific model vs chance), confirming our results that even this effect size is significant ($p = 0.02$).

Overall the previous comparison clearly demonstrates that on average person-specific models were more accurate than the classical approach, which assumes that the same neural mechanisms operate similarly in each person, a prediction captured in the nonspecific model. Moreover, for both tasks and all age groups, performance was better than random guesses for each trial.

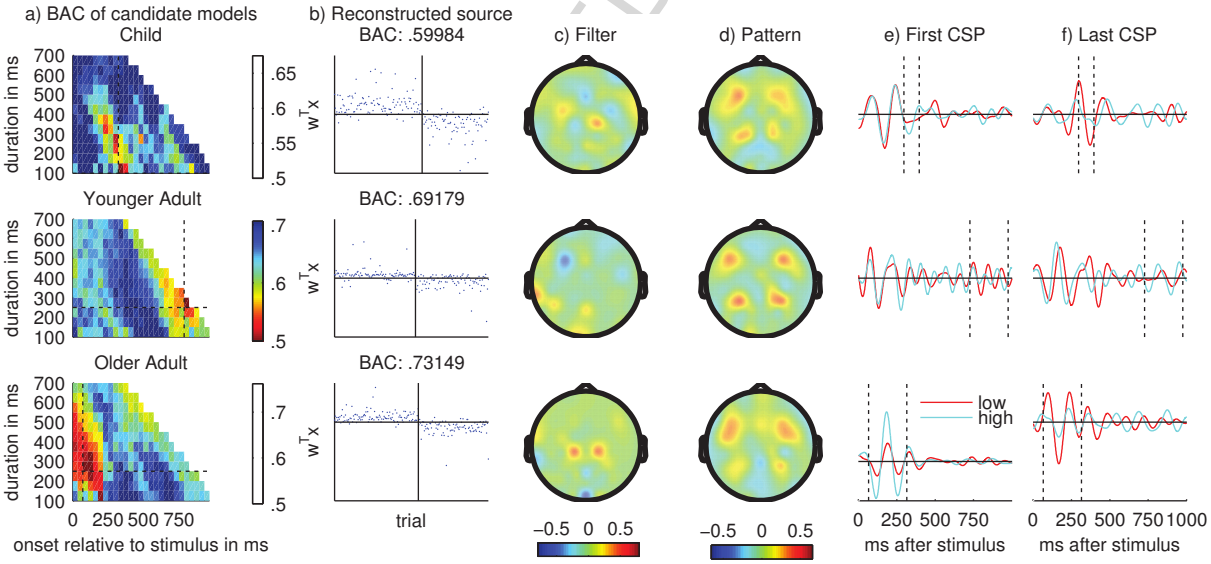
3.2. Person-specific results

Now that we have established that the person-specific models were more accurate than chance and the nonspecific model, the question is how to interpret the resulting estimated models both on an individual and on a group level. In the following we will describe how we interpreted the person-specific models and how we summarized the individual results on the group level.

In the supplement we show the person-specific results for all persons in both tasks. In Fig. 5 we show the



(a) Attentional focus



(b) WM load

Figure 5

Figure 5: **Person-specific results.** Examples of results for a) attentional focus and b) WM load prediction for single individuals, one from each age group. Column (a) shows the estimated BAC for the different candidate models. The x-axis describes the onset and the y-axis the duration of the corresponding time window. Colors refer to the estimated BAC of the respective candidate model, with hot colors indicating higher BAC and cold colors lower BAC. The crosshair depicts the location of the selected model. Column (b) shows the estimated source component for each trial as reconstructed by the best estimated model, including the BAC of the best model. The trials are sorted by their true class. The vertical line separates the classes. The horizontal line marks 0. For a perfect classifier the estimated source must be positive for all trials to the left and negative for all trials to the right of the vertical line. Column (c) shows the entries of the normalized filters. Column (d) shows the normalized pattern. Column (e) shows the mean time series for the first CSP filter for both classes. Column (f) shows the mean time series for the last CSP filter for both classes. The x-axis describes the time elapsed since the onset of the memory array. The vertical dotted lines indicate the selected time window. The horizontal line in columns e and f) marks 0.

results for the person with the most accurate estimated model in each age group and task. The properties of the selected person-specific models quantify different aspects of the observed data. Each candidate model was derived from the same basic class, but they vary with regard to exposure to certain spatio-temporal features of the observed brain responses. The onset and duration of time windows each model is exposed to precisely describe the temporal information used by a given model. Within the spatial domain, the resulting spatial filter and pattern coefficients reflect the topographical information exploited by the chosen model.

The BACs estimated for the different candidate models (column a in Fig. 5) map out a space of information content across different choices of time windows. They also quantify the uncertainty of the model selection procedure. If there is only one model with a high BAC, one can be relatively certain that the truly best model was selected. If there is almost no difference between the model with the highest BAC and several other models, however, model selection is mostly determined by random variation. We only report the BACs of the candidate models that have the same number of CSP filter pairs as the most accurate candidate model. This is motivated by the fact that we are mostly interested in the location of the time window. In addition, the number of CSP filter pairs was the least critical setting by which the candidate models differed (see Section 3.4). Furthermore, we report the following properties of the person-specific models: the BAC estimate as obtained by nested cross-validation, the classifier output (= reconstructed source) $\mathbf{w}^\top \mathbf{x}_t + c$ for each trial t sorted by true class membership as a visualization of the predictive behavior (both column [b]), the filter (column [c]) and the pattern (column [d]) to describe the spatial information that was employed by the model, the selected time window to quantify which time segment was employed for prediction, and the mean (over trials) time series of the first (column [e]) and last (column [f]) CSP component for each class as a visualization of the time course of the most discriminative projected components.

We will exemplarily describe the results for the younger adult in the attention task (second row in Fig. 5). The grid of candidate model BACs shows that candidate models employing time windows starting early

have a high BAC (column [a]), while none starting later than 200 ms achieves a comparable BAC. The reconstructed source shows the nice separability of the two conditions for this person, which is also reflected in a BAC of 0.8464 (column [b]). The filter has a high negative value at Pz, indicating that higher power at this electrode is evidence for the "Attention right" condition (column [c]). The pattern has very high positive values in parieto-occipital areas, indicating that if the decision function predicts "Attention Left", these areas show higher power (column [d]). The longest possible time window was selected (700 ms) and starts directly after stimulus onset. The CSP components both show the desired effect within the chosen time window: increased variance for one class and decreased variance for the other class. This effect is not present outside the selected time window (columns [e] and [f]).

3.3. Group results

Figs. 6 and 7 show the group results. The scatter plot of onsets and durations of the selected time windows (Figs. 6a and 7a) illustrate their between-persons variability. The mean (across persons) BAC for the different candidate models is shown in Figs. 6b and 7b. As for the person-level results, we are most interested in the location of the time window, so we only report the mean by onset and duration of the employed time window. To achieve this we averaged across the different numbers of CSP filter pairs within a person prior to averaging across all persons. The mean BAC for the different candidate models map out a space of mean information content across different choices of time windows.

Figs. 6c–e and Figs. 7c–e show three time series. The first expresses how often each time point was part of the selected model (Figs. 6c and 7c respectively). For each time point, we counted the number of persons for whom this time point was included in the person-specific model and then divided this by the total number of persons in this group. This is a way of visualizing the information per time point. However, as it only relies on the best estimated model it ignores the useful information of the candidate models that were not selected. Therefore, for a second time series we compressed the mean BAC of the candidate models in a similar fashion by calculating the mean BAC for every time point (Figs. 6d and 7d). We did this by averaging the mean BACs of all estimated models that contain the respective time point. For models that employ a very long-lasting time window, however, this could be misleading as the high performance may be driven by a very localized part. To remedy this, we also report the mean BAC for every time point but limited to the candidate models based on the shortest time window (100 ms) (Figs. 6e and 7e). We will refer to these results as coarse (Figs. 6d and 7d) and fine mean (Figs. 6e and 7e) BAC by time.

We do not report any averages of the person-specific filters and pattern, as they originate from different time windows and thus, their interpretation is difficult.

3.3.1. Attentional focus

Fig. 4a showed that mean BAC was highest for the older adults (0.6669), followed by the younger adults (0.6390), and lowest for the children (0.5538) for attentional focus. Fig. 6 shows the group results.

For the children, the employed time window of the best estimated model varied across the whole range of durations and onsets (Fig. 6a). This finding may indicate larger variability in the group of children compared

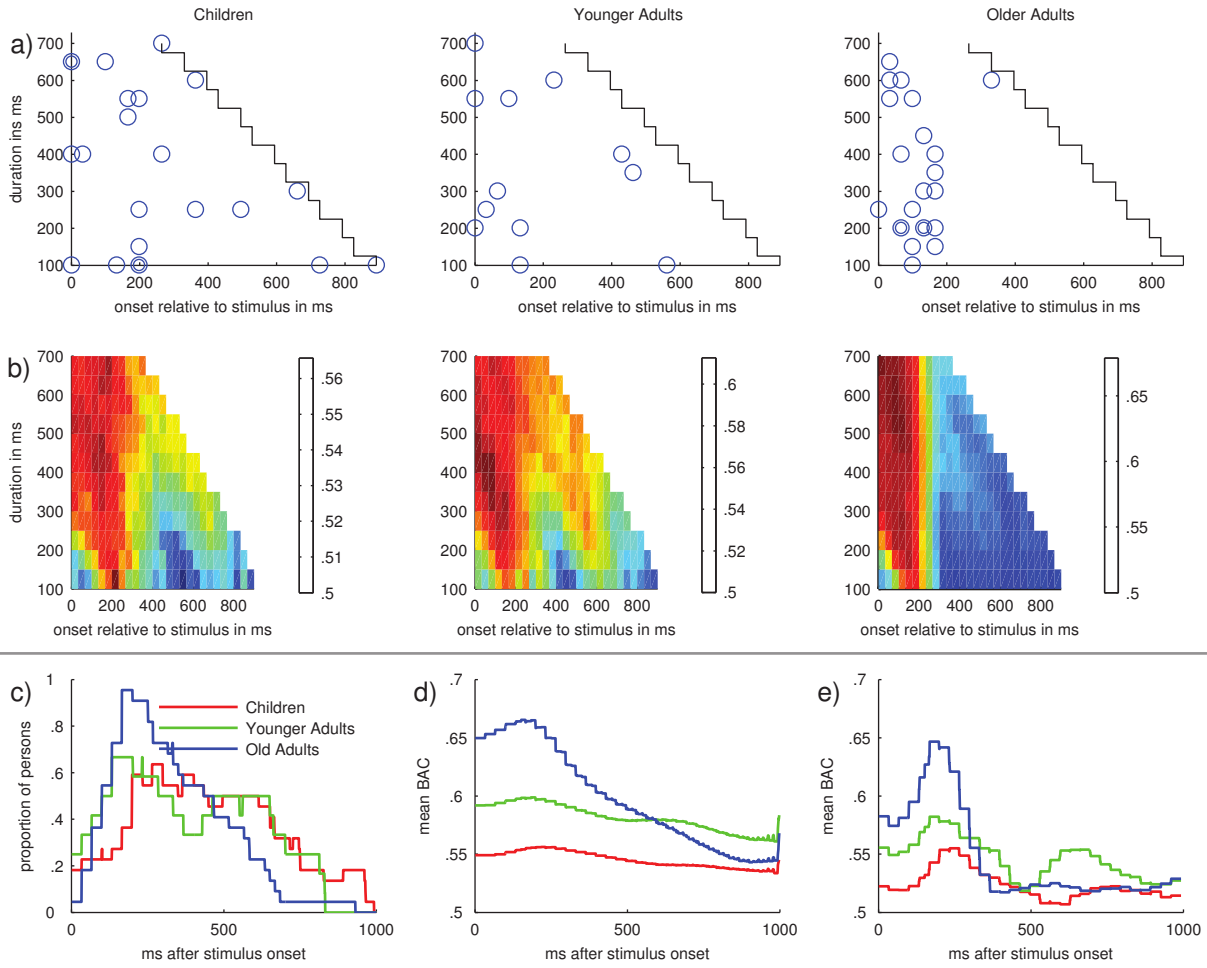


Figure 6: **Group results for the attentional foci prediction.** In row (a) and (b), each column shows the results for one age group. In row (a) each point represents the corresponding time window of one person-specific model. Double rings indicate that the corresponding time window was selected for two persons. The x-axis describes the onset and the y-axis the duration of the selected model. The step function encloses the region of possible time windows. In row (b) the mean BAC for the candidate models by onset (x-axis) and duration (y-axis) are shown. Colors refer to the estimated BAC of one candidate model, with hot colors referring to higher BAC and cold colors to lower BAC. Panel (c) shows for each time point the proportion of persons whose model used the respective time point, for each age group. Panel (d) shows the mean BAC by time using all candidate models (coarse mean BAC by time). Panel (e) shows the mean BAC by time using only the candidate models with a time window lasting 100 ms (fine mean BAC by time).

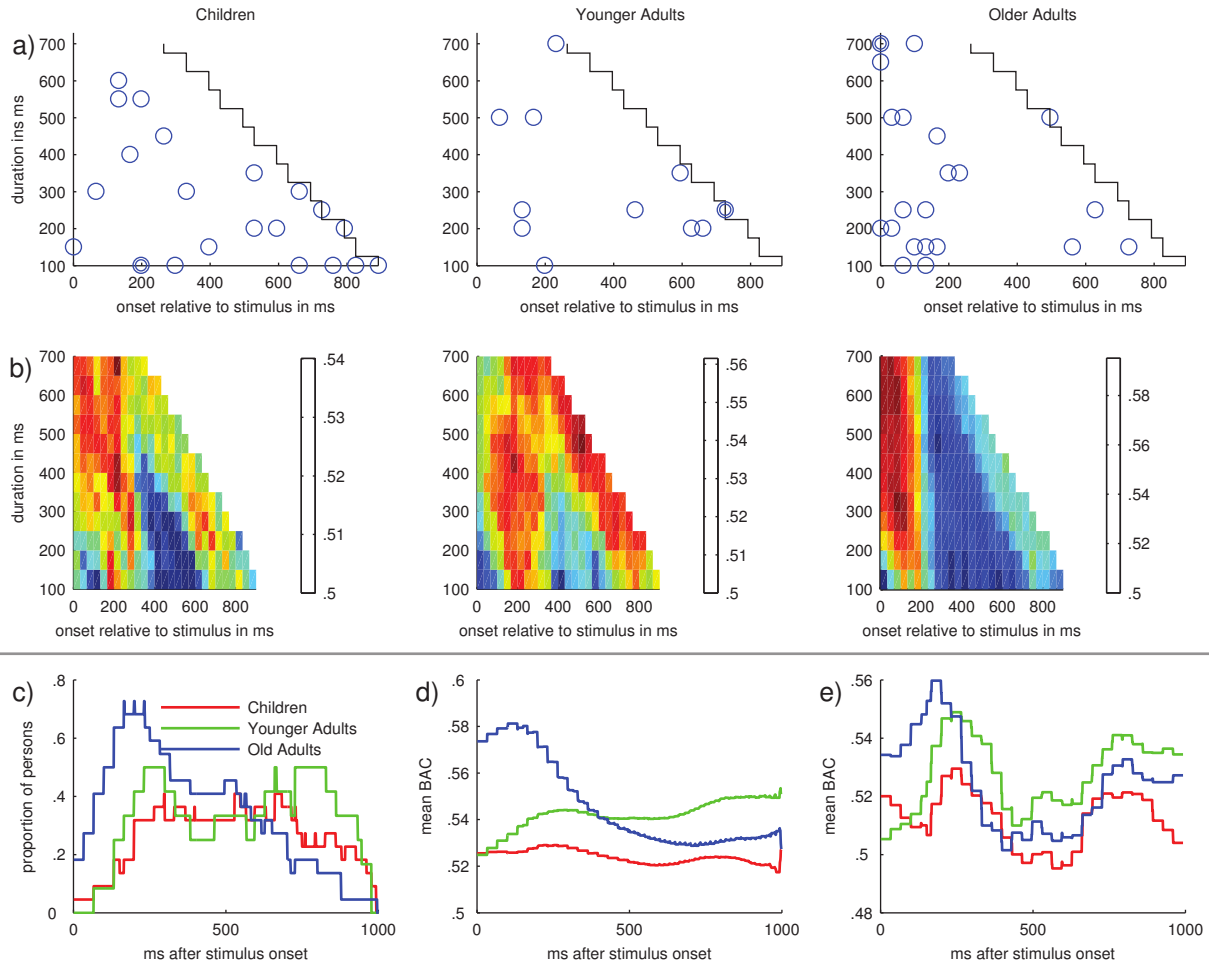


Figure 7: **Group results for the WM load prediction.** For a description see Fig. 6.

to the other age groups. However, given that the mean BAC was only slightly higher than chance level, this must be interpreted with caution.

For the majority of younger adults, models that start earlier than 250 ms after memory array onset were selected (Fig. 6a). For the remaining three younger adults, models starting later than 400 ms were picked. No model beginning later than 600 ms was selected. However, some of the early models contain data for almost the complete trial because of their duration (see supplement). The fine mean BAC by time plot (Fig. 6e) shows two distinctive peaks around 200 ms and around 700 ms for the younger adults. This is a consequence of the fact that in our group of younger adults, we observed two clusters: For one cluster, an early time window provided good discriminability, and for the other, a later time window discriminated well. This is reflected by the fact that the BAC for the candidate models peaked around the best model and was relatively low for the remaining candidate models for every younger adult (see supplement).

The distribution of time windows in the older adults is strongly shifted to the stimulus onset. For only one older adult a model starting later than 200 ms post-stimulus was selected, with a relatively low BAC of 0.5417 (Fig. 6a). The best estimated model contains data after 700 ms post-stimulus for only one person, while almost all person-specific models contain data around 200 ms post-stimulus. The coarse (Fig. 6d) and fine (Fig. 6e) mean BAC by time plots clearly show an early peak (at roughly 200 ms) followed by a sudden decline. This indicates that for almost all older adults, discriminability was best if very early time windows were employed. It is noteworthy that the accuracy of these models was superior to those of the younger adults (see Fig. 4a).

3.3.2. Working memory load

The mean BAC for WM load was highest for the the younger adults (0.5701), followed by the older adults (0.5565), and lowest for the children (0.5276) (see Fig. 4b). Fig. 7 shows the group results for the WM load prediction.

For the children, the employed time window of the best estimated model varied across the whole range of duration and onset (Fig. 7a). Again, this finding may indicate larger variability in the group of children compared to the other age groups. However, given that the mean BAC was only slightly higher than chance level, this finding must be treated with caution. For many of the younger adults and the older adults (6 of 12 younger adults and 18 of 22 older adults), models that started earlier than 250 ms after memory array onset were chosen (Fig. 7a). After a gap, there is another group of people with models that start after 450 ms (6 younger adults and 4 older adults). Note that the group of persons for whom early models were picked is much larger in the older than in the younger adults. Some of the early models contain data for almost the complete trial because of their duration.

Accordingly, the mean BAC for the candidate models shows two peaks for the younger and the older adults (Fig. 7b). One represents early models and the other late models. For the younger adults, both peaks are equally strong. For them, the coarse BAC by time plot (Fig. 7d) is therefore almost constant. Consequently, their fine BAC by time plot (Fig. 7e) shows an early and a late peak.

Despite the fact that there is a peak for late models, the coarse BAC by time plot (Fig. 7d) for the older

adults is qualitatively the same as for the attentional focus prediction: An early peak around 200 ms and then a sudden drop. The fine mean BAC by time reveals both peaks.

For both age groups, the reasons for the late and the early peak are less clear than for the attentional focus prediction. There are persons for whom only early models were accurate, persons for whom only late models were accurate, and persons for whom early and late models were accurate (see supplement).

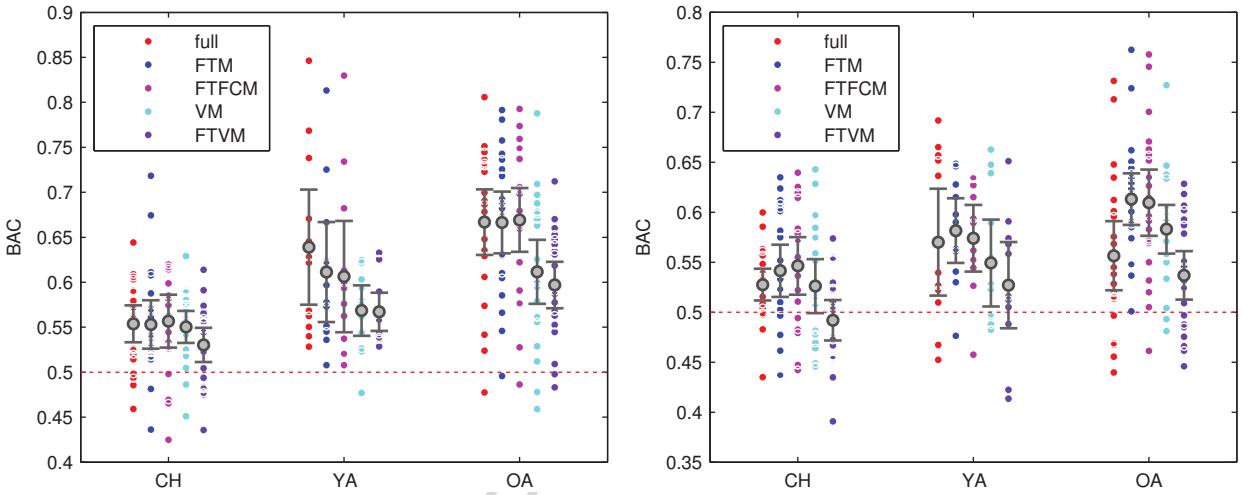
3.4. Performance comparison against simpler person-specific models

The set of candidate models that we proposed in order to obtain person-specific models is comparatively flexible (i.e., it adapts many parameters to the data) and uses sophisticated algorithms: model estimation entails parameter selection over time windows, a spatial projection, and a subsequent estimation of regularized regression weights. As a consequence results are hard to interpret. In the following, we report the results of an evaluation of the impact of reducing parameters to arrive at potentially simpler models. To this end, we gradually reduced the complexity of the candidate models. All the p-values that we report in this section were calculated using a two-sided paired t-test, as we hypothesized that the simpler model classes would be either better or worse then our original model class. At first sight it might seem unreasonable to hypothesize that a simpler model class leads to better predictive performance, as more complex model classes always lead to a better model fit. However, a better model fit does not necessarily lead to better predictive performance. This is due to the fact that a more flexible model class is more vulnerable to noise in the data. This is called overfitting in the literature (Bishop, 2007; Duda et al., 2001; Hastie et al., 2001).

In a first step, we abandoned the optimization across the time window by fixing it to a theory-driven estimate for each task (0–400 ms for attentional focus, 400–1000 ms for WM load). We call this model class the *fixed time model* in the following. By fixing the time windows we reduced the number of candidate models from 1220 (244 time windows by up to 5 components) to 5. For the attentional focus prediction, the mean BAC of the fixed time model was not significantly different from the mean BAC of the full model class for both the children ($p = 0.9605$) and the older adults ($p = 0.9265$). The mean BAC of the full model was slightly higher for the younger adults ($p = 0.1455$). The results for the older and younger adults are not surprising considering our previous analysis: For almost every older adult, a model with an early component was accurate. For the younger adults, time window optimization improved the results as there were two clusters: one for which an early, and one for which a late component is most predictive of the behavioral difference. For the children this again suggests that the observation of large variability might in fact be due to noise.

Regarding the WM load prediction, the mean BAC of the fixed time model was not significantly different from the mean BAC of the full model class for both the children ($p = 0.3465$) and the younger adults ($p = 0.5731$). For the older adults, the mean BAC of the fixed time model was higher than the mean BAC of the full model ($p = 0.0013$).

As a further step of model simplification, we abandoned the optimization of the number of CSP filter pairs. Instead, we picked the first three filter pairs. This was previously found to yield good predictive performance (Blankertz et al., 2008). We refer to this model as the *fixed time fixed CSP model*. Across all age groups and



(a) Attentional focus (b) WM load

Figure 8: **Performance comparison.** Mean BAC and CI for the five different model classes for each of the three age groups within the attentional focus (a) and WM load prediction tasks (b). Each dot describes the BAC of one person. The box plots denote the respective mean values and 95% CIs. The red horizontal line describes the BAC as expected by the null hypothesis. "CH" stands for children, "YA" for younger adults, and "OA" for older adults. "FTM" stands for "fixed time model," "FTFCM" for "fixed time fixed CSP model,", "VM" for "variance model," and "FTVM" for "fixed time variance model" (see text for details).

tasks, the additional restriction of the model space did not significantly influence the accuracy in comparison to the fixed time model (attentional focus: children: $p = 0.6075$, younger adults: $p = 0.4311$, and older adults: $p = 0.5450$; WM load: children: $p = 0.5600$, younger adults: $p = 0.4184$, and older adults: $p = 0.6980$). This is further evidence that three filters represent a reasonable choice.

To judge if CSP is a reasonable feature extraction method for the present data, we introduced two more model classes that do not rely on CSP features but directly operate on the variance of each channel. The two classes differed from each other in the following: In one class, we optimized the time window as before. We call this model class *variance model* in the following. We compared this class against the full model. For the other class, we fixed the time window to our theory-driven estimates (0–400 ms for attention, 400–1000 ms for load). We call this model class *fixed time variance model*. We compared this class against the fixed time model. For the attentional focus prediction the variance model was less accurate than the full model for both the younger and the older adults ($p = 0.0256$ and $p = 0.0004$). For the children there was no significant difference ($p = 0.7986$). For the WM load prediction there was no significant difference between the variance model and the full model for any age group (children: $p = 0.9193$, younger adults: $p = 0.2728$, and older adults: $p = 0.1355$). The fixed time model was more accurate than the fixed time variance model for both tasks and all age groups (attentional focus: children: $p = 0.0902$, younger adults: $p = 0.0533$, and older adults: $p < 0.0001$; WM load: children: $p = 0.0019$, younger adults: $p = 0.0382$, and older adults: $p = 0.0007$). Thus, for most comparisons, CSP features led to more accurate models. However, especially for WM load prediction for the older adults the combination of CSP and time selection seems to lead to overfitting.

4. Discussion

The aim of the current study was to test and evaluate a multivariate pattern classification approach on EEG data from a lifespan sample. The study was motivated by the observation of increased behavioral and neuronal heterogeneity in both children and older adults (Lindenberger et al., 2013; Nagel et al., 2009; Werkle-Bergner et al., 2012), challenging the validity of group-based analyses as well as between-group comparisons. The results of the present study demonstrate the potential of multivariate pattern classification approaches (e.g., Bishop, 2007; Blankertz et al., 2011; Parra et al., 2005) for the derivation of idiographic models (Nesselroade et al., 2007) in age-comparative EEG studies. WM load as well as the focus of spatial attention could be discriminated reliably in all three age groups based on rhythmic neural activity in the alpha frequency range. In all three age groups, the BAC of person-specific models provided a significantly above-chance classification for the prediction of both the attentional focus and WM load. Also, across age groups and for both prediction targets, the person-specific models were more accurate than a theory-driven nonspecific model disregarding individual differences, with the exception of the children's model for the prediction of WM load. Hence, the present framework demonstrates the feasibility of deriving person-specific models that capture the idiosyncratic signatures of neural mechanisms related to changing attentional foci and WM load as reflected in varying spatio-temporal properties of the measured EEG signals.

Usually, the spatio-temporal differences in the neural response between persons are assumed to reflect noise (e.g., Luck, 2005). However, our approach shows that individually-tailored models do not just fit noise, as exemplified by above-chance BAC derived from nested cross-validation. Rather, in terms of prediction accuracy, they outperform conventional nonspecific models that assume operation of the same neurobehavioral mechanisms in each participant (e.g., Danziger, 1990). This finding underscores the idiosyncrasy of neural responses underlying similar overt behaviors, and calls for further studies to investigate how this is related to interindividual differences both on the neural level (e.g., in terms of differences in brain structure; cf. Breakspear et al., 2010; Deco et al., 2012; Valdes-Hernandez et al., 2010) and on the behavioral level (e.g., in terms of differences in strategy use; cf. Corbin and Marquer, 2009).

One of the major challenges when dealing with person-specific models is how to extract and aggregate the person-specific information across individuals in order to allow for group comparisons. Here, we suggest summarizing person-specific models with regard to their different properties. These properties refer to the timing of the processes (here: time windows with a certain onset and duration), the reliability of classification within the person (BAC), and the topographical distribution of classification information (i.e., filters and patterns). Summarized in this way, the information can be taken to visualize the variability between persons and to compare the distribution of model properties across groups. Some quantities might even be subject to traditional statistical inference. For example, filters and patterns are described by a vector and could simply be averaged. In this way, the data can be brought back into commonly reported topographical space and compared across groups. However, one must be aware that filters and patterns originate from very different time windows and thus the meaning of averages may be questioned. Summarizing person-specific models on the group level poses a serious challenge both on a conceptual and on a methodological dimension. We see our approach as a first step in this direction.

In terms of classification accuracy, differences between the age groups were observed. With regard to the attentional focus prediction, the BAC was highest for older adults, followed by younger adults, and lowest for children. Concerning the WM load classification, the highest BACs were obtained for the younger adults, followed by the older adults, with the lowest values again for children. The reduced overall classification performance in the group of children, which also holds true for the simpler alternative models, underlines that single-trial classification failed for a considerable proportion of children and thus requires cautious interpretation of the results. We tentatively suggest that the discrimination of cognitive states based on EEG oscillations is more demanding in children due to fluctuations in sustained attention (cf., Fandakova et al., 2014; Sander et al., 2011). Fluctuating attentional states are known to affect performance in change–detection tasks considerably (Rouder et al., 2008). Interestingly, and in contrast to our expectations, classification performance was not reduced in older adults, and was even higher, although not significantly, than in younger adults with regard to the classification of the attentional focus. Comparing the most accurate model classes for the WM load prediction between age groups (fixed time fixed CSP model for the children, fixed time model for the younger adults and older adults) the older adults are again the ones with the highest BAC. This observation is in line with the observation that it is not potential group heterogeneity per se that impairs discrimination performance, but rather variability within a person that might negatively impact classification

(e.g., Krauledat et al., 2008).

One typical challenge that developmental researchers face is dealing with age differences in the timing of (cognitive) processing (Kail and Salthouse, 1994; Salthouse, 1996). Here, such idiosyncrasies are considered by deriving a set of candidate models that differ with regard to the temporal dimensions, (i.e., by varying the onset and duration of observed time windows). The present results demonstrate that the most discriminative neural responses can be observed at different onsets and with variable duration, suggesting interindividual differences in processing speed and reliability (Li et al., 2004). Going beyond simple differences in processing speed for the supposedly same cognitive process, the observed differences may also indicate differences in cognitive processing strategies. In the group of children, time windows of the best models varied across the whole range of possible onsets and durations for both prediction tasks. Together with the low BAC in this age group, this pattern underlines our interpretation of unsustained task processing in the child group rendering data interpretation challenging.

In the group of younger adults, both prediction tasks revealed interindividual differences in the onset and duration of time windows that discriminated optimally between conditions. In most younger adults, a model was selected that starts earlier than 250 ms after memory array onset. This observation is in line with early effects of attention, closely following cue or stimulus onset (e.g., Broadbent, 1958; Hackley et al., 1990; Hillyard et al., 1998). However, the fine mean BAC by time plot (see Fig. 6e) revealed two peaks in this group, one early peak at around 200 ms and a later one at around 700 ms. The later peak indicates some individuals that seem to shift their attentional focus rather late and thus, may rely on late selection processes (Deutsch and Deutsch, 1963). In older adults, we found a strong shift of the time windows towards stimulus onset. In all but one participant, the best model's onset was found before 200 ms. It has been debated for decades whether selective attention operates rather early (e.g., Broadbent, 1958) or late (Deutsch and Deutsch, 1963) within the temporal sequence of information processing. Most likely, the reliance on early versus late selection mechanisms differs individually. In addition, early versus late selection has been suggested to be a function of cognitive load, with low load conditions leading to rather early, and high load conditions leading to rather late selection processes (Lavie, 2005). Interestingly, current results in cognitive aging research underline that cognitive load is dependent on the individual processing capacity, which differs between age groups (e.g., Nagel et al., 2009; Werkle-Bergner et al., 2012; Schneider-Garces et al., 2010). Accordingly, younger adults' low load conditions may equate high load conditions for older adults, leading in turn to a shift to early selection mechanisms in conditions in which younger adults can still recruit late selection processes (but see Velanova et al., 2007, for an alternative view). In line with this interpretation, we also observed a strong peak of the classification accuracy in an early time window for the older adults in the WM load prediction task, but an early and a late peak in the group of younger adults. These results reveal that interindividual differences in timing are relevant for the interpretation of neural responses in the context of current theories of selective attention.

The discussion of the results so far was solely based on the full model. However, from a classification perspective, the full model was only as accurate as a simpler model class using fixed time windows (fixed time model) in conjunction with CSP and LDA. These time windows were derived from the literature and

were constrained to an early time window for the classification of the attentional focus and to a later time window for the classification of the WM load. Using these fixed windows, with exception of the WM load prediction in the older adults, we almost achieved the same BACs as with time window optimization. Hence, two models, the full model and the fixed time model, do not differ significantly in predictive accuracy. In other words, the hypothesis that their performance on a new observation is equal can not be rejected. It is generally taken for granted that more parsimonious explanations for a set of observations are preferable under comparable goodness-of-fit (Akaike, 1998). In our case the more parsimonious explanation corresponds to the fixed time model, as it is a special case of the full model. But, since we relied on cross-validation to estimate the predictive accuracy, we implicitly accounted for the unequal complexity between the models in terms of numbers of free parameters. In fact, one can show for likelihood-based models that cross-validation is asymptotically equivalent to the Akaike information criterion that penalizes model fit for the number of parameters (Stone, 1977). Hence, unless model fit was evaluated directly for new observations, choosing the simpler model may be a good heuristic – or it may not. Insofar, we allow ourselves to draw conclusions from the two hypotheses in head-to-head competition. In other words, the full model provides evidence for interindividual variability in brain responses, which would have remained covered by the fixed time model analysis alone. Hence, the person-specific models, although not superior with regard to prediction accuracy, provide additional insights about reliable idiosyncrasies in brain–behavior mappings useful for further theory-building. Interestingly, for older adults the BAC achieved by our time-optimizing model was even worse than that of the fixed time model. One possible explanation may be that for the WM load prediction, the discriminating signal is too weak so that our complex model class, containing many candidate models, resulted in overfitting of the data. That is, the model selection process is disturbed by the noise in the data such that a theory-driven model selection leads to more accurate predictions. From a theoretical perspective one could argue that the discrimination between two load conditions is particularly challenging in the group of older adults. The reason being that the different load conditions may indeed be less distinct due to lower capacity limits.

While standard EEG analysis approaches assume the same model for every person, we made the assumption that a different model is required for every person. The two approaches represent the extremes of a continuum. It would be very interesting to systematically explore this continuum. That is, to examine how much individualization of the models is necessary. We envision two different approaches to reach this goal. The first would start with one model for all participants and recursively split the participants into groups with the same model. The participants are only split further if a significant gain in prediction accuracy can be obtained. Brandmaier et al. (2013) already developed this framework for Structural Equation Models. An alternative approach would be a hierarchical model that penalizes the person-specific models for a derivation from the group model. The best amount of penalty could be estimated and would be informative regarding the amount of individualization necessary for a given task. Both approaches are accompanied by their own methodological and computational challenges.

To derive the person-specific models we employed the combination of CSP, LDA, and nested cross-validation. This setup is very popular for building BCIs based on rhythmic neural activity (e.g., Blankertz

et al., 2008, 2011; Ramoser et al., 2000). Our work is distinct from typical BCI studies as we did not aim to construct a real-time classification system but rather used the same methods for post-hoc data analysis. Therefore, in contrast to typical BCI studies, we were able to improve the data quality prior to applying our classification system. This included estimating the individual alpha peak frequency for every person and using ICA-based artifact correction. We overcame the technical challenge that data corrected by ICA are not of full rank, which is necessary for CSP, by applying PCA (Blankertz et al., 2008). We believe that this preprocessing was crucial for the success of this study because of the commonly reduced amount of data in developmental studies. EEG data quality in children and older adult groups is often reduced due to more oculo-motor, muscle, or movement artifacts (Picton et al., 2000), leaving fewer trials per condition. In addition, reduced attentional capacities of these groups (McAvinue et al., 2012) call for shorter testing sessions than in younger adults. Overall, the use of analysis techniques derived from pattern classification and BCI research is a valuable tool to overcome some of the limitations in terms of data quality typically encountered in age-comparative neuroimaging studies.

For the attentional focus prediction, the experimental condition remained the same for blocks of 30 trails each. If slowly varying background activity that could be exploited by the classification algorithm were present in a blocked experimental design, using k-fold cross-validation might overestimate the performance of the classifier. A general remedy would be to use leave-one-block-out cross-validation instead (Lemm et al., 2011).

In sum, the present study evaluated a multivariate pattern classification approach for the analysis of idiographic models in developmental cognitive neuroscience studies by demonstrating its general feasibility in three age groups. In our view, an improved consideration of interindividual differences in brain–behavior mappings is one of the upcoming challenges for cognitive neuroscience research in developmental samples (cf., Molenaar, 2004; Molenaar and Campbell, 2009; Nesselroade et al., 2007; Voelkle et al., 2014). In this context, developing analytic approaches that account for idiosyncrasies in the mappings between neural and cognitive processes is necessary for further progress in lifespan cognitive neuroscience. The present study can be conceived as a further step in this direction.

5. Acknowledgments

This study was conducted within the projects, “Cognitive and Neuronal Dynamics of Memory across the Lifespan (CONMEM)” and “Formal Methods in Lifespan Psychology” at the Center for Lifespan Psychology, Max Planck Institute for Human Development, Berlin, Germany. The Max Planck Society and the German Research Foundation (DFG, HE 3347/3-1) financially supported the research. The funding sources were not involved in any decision regarding this study. Julia Delius provided highly valuable editorial assistance. The study was conducted in partial fulfillment of the doctoral dissertation of JDK.

6. References

Akaike, H., 1998. Information theory and an extension of the maximum likelihood principle. In: Parzen, E., Tanabe, K., Kitagawa, G. (Eds.), Selected Papers of Hirotugu Akaike. Springer Series in Statistics. Springer New York, pp. 199–213.

Astle, D. E., Scerif, G., 2011. Interactions between attention and visual short-term memory (VSTM): What can be learnt from individual and developmental differences? Neuropsychologia 49 (6), 1435–1445.

Bäckman, L., Nyberg, L., Lindenberger, U., Li, S.-C., Farde, L., 2006. The correlative triad among aging, dopamine, and cognition: current status and future prospects. Neurosci. Biobehav. Rev. 30 (6), 791–807. URL http://dx.doi.org/10.1016/j.neubiorev.2006.06.005

Bahramisharif, A., van Gerven, M., Heskes, T., Jensen, O., Apr 2010. Covert attention allows for continuous control of brain-computer interfaces. Eur. J. Neurosci. 31 (8), 1501–1508. URL http://dx.doi.org/10.1111/j.1460-9568.2010.07174.x

Bießmann, F., Daehne, S., Meinecke, F., Goergen, K., Blankertz, B., Haufe, S., 2012. On the interpretability of linear multivariate neuroimaging analyses: Filters, patterns and their relationship. In: Proceedings of the 2nd NIPS Workshop on Machine Learning and Interpretation in Neuroimaging.

Bishop, C. M., 2007. Pattern Recognition and Machine Learning (Information Science and Statistics), 1st Edition. Springer.

Blankertz, B., Lemm, S., Treder, M., Haufe, S., Müller, K.-R., 2011. Single-trial analysis and classification of ERP components — a tutorial. Neuroimage, 814–825.

Blankertz, B., Tangermann, M., Vidaurre, C., Dickhaus, T., Sannelli, C., Popescu, F., Fazli, S., Danóczy, M., Curio, G., Müller, K.-R., 2010. Detecting mental states by machine learning techniques: The berlin brain–computer interface. In: Graimann, B., Pfurtscheller, G., Allison, B. (Eds.), Brain-Computer Interfaces. The Frontiers Collection. Springer, Berlin, Heidelberg, pp. 113–135.

Blankertz, B., Tomioka, R., Lemm, S., Kawanabe, M., Muller, K.-R., 2008. Optimizing spatial filters for robust eeg single-trial analysis. Signal Processing Magazine, IEEE 25 (1), 41–56.

Boker, S. M., Molenaar, P., Nesselroade, J. R., 2009. Issues in intraindividual variability: individual differences in equilibria and dynamics over multiple time scales. Psychol. Aging 24 (4), 858.

Brandmaier, A. M., von Oertzen, T., McArdle, J. J., Lindenberger, U., 2013. Structural equation model trees. Psychological Methods 18 (1), 71.

Breakspear, M., Jirsa, V., Deco, G., Sep 2010. Computational models of the brain: from structure to function. Neuroimage 52 (3), 727–730.
URL http://dx.doi.org/10.1016/j.neuroimage.2010.05.061

Broadbent, D. E., 1958. Perception and communication. Pergamon Press, Oxford, UK.

Brodersen, K. H., Ong, C. S., Stepan, K. E., Buhmann, J. M., 2010. The balanced accuracy and its posterior distribution. In: International Conference on Pattern Recognition.

Corbin, L., Marquer, J., Jun 2009. Individual differences in sternberg's memory scanning task. Acta Psychol (Amst) 131 (2), 153–162.
URL http://dx.doi.org/10.1016/j.actpsy.2009.04.001

Danziger, K., 1990. Constructing the subject: Historical origins of psychological research. Cambridge University Press, Cambridge, UK.

Deco, G., Senden, M., Jirsa, V., 2012. How anatomy shapes dynamics: a semi-analytical study of the brain at rest by a simple spin model. Front Comput Neurosci 6, 68.
URL http://dx.doi.org/10.3389/fncom.2012.00068

Deutsch, J. A., Deutsch, D., 1963. Attention: some theoretical considerations. Psychol. Rev. 70 (1), 80.
URL http://psycnet.apa.org/journals/rev/70/1/80/

Doppelmayr, M. M., Klimesch, W., Pachinger, T., Ripper, B., 1998. The functional significance of absolute power with respect to event-related desynchronization. Brain Topogr. 11 (2), 133–140.

Duda, R. O., Hart, P. E., Stork, D. G., 2001. Pattern Classification, 2nd Edition. Wiley, New York.

Fandakova, Y., Sander, M. C., Werkle-Bergner, M., Shing, Y. L., Mar 2014. Age differences in short-term memory binding are related to working memory performance across the lifespan. Psychol. Aging 29 (1), 140–149.
URL http://dx.doi.org/10.1037/a0035347

Fazli, S., Popescu, F., Danóczy, M., Blankertz, B., Müller, K.-R., Grozea, C., 2009. Subject-independent mental state classification in single trials. Neural Networks 22 (9), 1305 – 1312.
URL http://www.sciencedirect.com/science/article/pii/S0893608009001063

Freunberger, R., Höller, Y., Griesmayr, B., Gruber, W., Sauseng, P., Klimesch, W., May 2008. Functional similarities between the p1 component and alpha oscillations. Eur. J. Neurosci. 27 (9), 2330–2340.
URL http://dx.doi.org/10.1111/j.1460-9568.2008.06190.x

Freunberger, R., Werkle-Bergner, M., Griesmayr, B., Lindenberger, U., Klimesch, W., Feb 2011. Brain oscillatory correlates of working memory constraints. Brain Res. 1375, 93–102.
URL http://dx.doi.org/10.1016/j.brainres.2010.12.048

Friedman, J., 1989. Regularized discriminant analysis. J. Am. Stat. Assoc. 84 (405), 165–175.

Grimault, S., Robitaille, N., Grova, C., Lina, J.-M., Dubarry, A.-S., Jolicœur, P., 2009. Oscillatory activity in parietal and dorsolateral prefrontal cortex during retention in visual short-term memory: Additive effects of spatial attention and memory load. Human brain mapping 30 (10), 3378–3392.

Hackley, S. A., Woldorff, M. G., Hillyard, S. A., 1990. Cross-modal selective attention effects on retinal, myogenic, brainstem, and cerebral evoked potentials. Psychophysiology 27 (2), 195–208.

Hastie, T., Tibshirani, R., Friedman, J. J. H., 2001. The elements of statistical learning. Vol. 1. Springer New York.

Hayes, K. J., 1953. The backward curve: a method for the study of learning. Psychol. Rev. 60 (4), 269.

Hillyard, S. A., Vogel, E. K., Luck, S. J., Hillyard, S. A., Vogel, E. K., Luck, S. J., 1998. Sensory gain control (amplification) as a mechanism of selective attention: electrophysiological and neuroimaging evidence. Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences 353 (1373), 1257–1270.
URL http://rstb.royalsocietypublishing.org/content/353/1373/1257

Jensen, O., Gelfand, J., Kounios, J., Lisman, J. E., Aug 2002. Oscillations in the alpha band (9-12 hz) increase with memory load during retention in a short-term memory task. Cereb. Cortex 12 (8), 877–882.

Jung, T., Makeig, S., Humphries, C., Lee, T., McKeown, M., Iragui, V., Sejnowski, T., 2000. Removing electroencephalographic artifacts by blind source separation. Psychophysiology 37, 163–178.

Kail, R., Salthouse, T. A., 1994. Processing speed as a mental capacity. Acta psychologica 86 (2), 199–225.

Kelly, S. P., Lalor, E. C., Reilly, R. B., Foxe, J. J., Jun 2005. Visual spatial attention tracking using high-density ssvep data for independent brain-computer communication. IEEE Transactions on Neural Systems and Rehabilitation 13 (2), 172–178.
URL http://dx.doi.org/10.1109/TNSRE.2005.847369

Kelly, S. P., Lalor, E. C., Reilly, R. B., Foxe, J. J., Jun 2006. Increases in alpha oscillatory power reflect an active retinotopic mechanism for distracter suppression during sustained visuospatial attention. J. Neurophysiol. 95 (6), 3844–3851.
URL http://dx.doi.org/10.1152/jn.01234.2005

Klimesch, W., 1999. {EEG} alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. Brain Research Reviews 29 (2–3), 169 – 195.
URL http://dx.doi.org/10.1016/S0165-0173(98)00056-3

Kohavi, R., et al., 1995. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: IJCAI. Vol. 14. pp. 1137–1145.

Krauledat, M., Tangermann, M., Blankertz, B., Müller, K.-R., 2008. Towards zero training for brain-computer interfacing. PLoS One 3 (8).
URL http://dx.doi.org/10.1371/journal.pone.0002967

Kray, J., Lindenberger, U., 2000. Adult age differences in task switching. Psychol. Aging 15 (1), 126–147.

Kriegeskorte, N., Simmons, W. K., Bellgowan, P. S., Baker, C. I., 2009. Circular analysis in systems neuroscience: the dangers of double dipping. Nat. Neurosci. 12 (5), 535–540.

Lavie, N., Feb 2005. Distracted and confused?: selective attention under load. Trends in Cognitive Sciences 9 (2), 75–82.
URL http://dx.doi.org/10.1016/j.tics.2004.12.004

Ledoit, O., Wolf, M., 2004. A well-conditioned estimator for large-dimensional covariance matrices. Journal of Multivariate Analysis, 365–411.

Lemm, S., Blankertz, B., Dickhaus, T., Müller, K.-R., 2011. Introduction to machine learning for brain imaging. Neuroimage 56 (2), 387–399.

Li, S.-C., Huxhold, O., Schmiedek, F., 2004. Aging and attenuated processing robustness. Gerontology 50 (1), 28–34.
URL http://www.karger.com/doi/10.1159/000074386

Lindenberger, U., Burzynska, A., Nagel, I. E., 2013. Heterogeneity in frontal-lobe aging. In: Stuss, D. T., Knight, R. T. (Eds.), Principles of frontal lobe function. Vol. 2. Oxford University Press, New York, pp. 609–627.

Luck, S. J., 2005. An Introduction to the Event-Related Potential Technique. Cambridge, MA: MIT Press.

Luck, S. J., Vogel, E. K., Nov 1997. The capacity of visual working memory for features and conjunctions. Nature 390 (6657), 279–281.
URL http://dx.doi.org/10.1038/36846

McAvinue, L. P., Habekost, T., Johnson, K. A., Kyllingsbæk, S., Vangkilde, S., Bundesen, C., Robertson, I. H., 2012. Sustained attention, attentional selectivity, and attentional capacity across the lifespan. Attention, Perception, & Psychophysics 74 (8), 1570–1582.

Molenaar, P. C., 2004. A manifesto on psychology as idiographic science: Bringing the person back into scientific psychology, this time forever. Measurement 2 (4), 201–218.

Molenaar, P. C., Campbell, C. G., 2009. The new person-specific paradigm in psychology. Current Directions in Psychological Science 18 (2), 112–117.

Müller-Gerking, J., Pfurtscheller, G., Flyvbjerg, H., 1999. Designing optimal spatial filters for single-trial eeg classification in a movement task. Clin. Neurophysiol. 110 (5), 787–798.

Nagel, I. E., Preuschhof, C., Li, S.-C., Nyberg, L., Bäckman, L., Lindenberger, U., Heekeren, H. R., 2009. Performance level modulates adult age differences in brain activation during spatial working memory. Proceedings of the National Academy of Sciences 106 (52), 22552–22557.

Nesselroade, J., Gerstorf, D., Hardy, S., Ram, N., 2007. Focus article: Idiographic filters for psychological constructs. Measurement: Interdisciplinary Research & Perspective 5 (4), 217–235.

Nettleton, D., Doerge, R. W., Mar. 2000. Accounting for variability in the use of permutation testing to detect quantitative trait loci. Biometrics 56 (1), 52–58.
URL http://www.jstor.org/stable/2677102

Norman, K. A., Polyn, S. M., Detre, G. J., Haxby, J. V., Sep 2006. Beyond mind-reading: multi-voxel pattern analysis of fmri data. Trends Cogn Sci 10 (9), 424–430.
URL http://dx.doi.org/10.1016/j.tics.2006.07.005

Obleser, J., Wöstmann, M., Hellbernd, N., Wilsch, A., Maess, B., Sep 2012. Adverse listening conditions and memory load drive a common ? oscillatory network. J. Neurosci. 32 (36), 12376–12383.
URL http://dx.doi.org/10.1523/JNEUROSCI.4908-11.2012

Oostenveld, R., Fries, P., Maris, E., Schoffelen, J.-M., 2010. Fieldtrip: open source software for advanced analysis of meg, eeg, and invasive electrophysiological data. Computational intelligence and neuroscience 2011.

Parra, L. C., Spence, C. D., Gerson, A. D., Sajda, P., 2005. Recipes for the linear analysis of eeg. Neuroimage 28, 326–341.

Pfurtscheller, G., Aranibar, A., Jun 1977. Event-related cortical desynchronization detected by power measurements of scalp eeg. Electroencephalogr. Clin. Neurophysiol. 42 (6), 817–826.

Philiastides, M. G., Sajda, P., 2006. Temporal characterization of the neural correlates of perceptual decision making in the human brain. Cereb. Cortex 16 (4), 509–518.

Picton, T., Bentin, S., Berg, P., Donchin, E., Hillyard, S., Johnson, R., Miller, G., Ritter, W., Ruchkin, D., Rugg, M., et al., 2000. Guidelines for using human event-related potentials to study cognition: recording standards and publication criteria. Psychophysiology 37 (02), 127–152.

Ramoser, H., Müller-Gerking, J., Pfurtscheller, G., 2000. Optimal spatial filtering of single trial eeg during imagined hand movement. Rehabilitation Engineering, IEEE Transactions on 8 (4), 441–446.

Raz, N., Lindenberger, U., Rodrigue, K. M., Kennedy, K. M., Head, D., Williamson, A., Dahle, C., Gerstorf, D., Acker, J. D., Nov 2005. Regional brain changes in aging healthy adults: general trends, individual differences and modifiers. Cereb. Cortex 15 (11), 1676–1689.
URL http://dx.doi.org/10.1093/cercor/bhi044

Rouder, J. N., Morey, R. D., Cowan, N., Zwilling, C. E., Morey, C. C., Pratte, M. S., 2008. An assessment of fixed-capacity models of visual working memory. Proceedings of the National Academy of Sciences 105 (16), 5975–5979.
URL http://www.pnas.org/content/105/16/5975.abstract

Roux, F., Wibral, M., Mohr, H. M., Singer, W., Uhlhaas, P. J., Sep 2012. Gamma-band activity in human prefrontal cortex codes for the number of relevant items maintained in working memory. The Journal of Neuroscience 32 (36), 12411–12420.
URL http://dx.doi.org/10.1523/JNEUROSCI.0421-12.2012

Royston, P., 1995. Algorithm as r94. Applied Statistics 44 (4).

Salthouse, T. A., 1996. The processing-speed theory of adult age differences in cognition. Psychological review 103 (3), 403.

Sander, M. C., Lindenberger, U., Werkle-Bergner, M., Oct 2012a. Lifespan age differences in working memory: a two-component framework. Neurosci. Biobehav. Rev. 36 (9), 2007–2033.
URL http://dx.doi.org/10.1016/j.neubiorev.2012.06.004

Sander, M. C., Werkle-Bergner, M., Lindenberger, U., Dec 2011. Contralateral delay activity reveals life-span age differences in top-down modulation of working memory contents. Cereb. Cortex 21 (12), 2809–2819.
URL http://dx.doi.org/10.1093/cercor/bhr076

Sander, M. C., Werkle-Bergner, M., Lindenberger, U., 2012b. Amplitude modulations and inter-trial phase stability of alpha-oscillations differentially reflect working memory constraints across the lifespan. Neuroimage 59, 646–654.

Sauseng, P., Klimesch, W., Heise, K., Gruber, W., Holz, E., Karim, A., Glennon, M., Gerloff, C., Birbaumer, N., Hummel, F., 2009. Brain oscillatory substrates of visual short-term memory capacity. Curr. Biol. 19, 1846–1852.

Schneider-Garces, N. J., Gordon, B. A., Brumback-Peltz, C. R., Shin, E., Lee, Y., Sutton, B. P., Maclin, E. L., Gratton, G., Fabiani, M., Apr 2010. Span, crunch, and beyond: working memory capacity and the aging brain. J. Cogn. Neurosci. 22 (4), 655–669.
URL http://dx.doi.org/10.1162/jocn.2009.21230

Siegler, R. S., 1987. The perils of averaging data over strategies: An example from children's addition. J. Exp. Psychol. Gen. 116 (3), 250.

Stone, M., 1974. Cross-validatory choice and assessment of statistical predictions. Journal of the Royal Statistical Society. Series B (Methodological), 111–147.

Stone, M., 1977. An asymptotic equivalence of choice of model by cross-validation and akaike's criterion. Journal of the Royal Statistical Society. Series B (Methodological), 44–47.

Vaden, R. J., Hutcheson, N. L., McCollum, L. A., Kentros, J., Visscher, K. M., Nov 2012. Older adults, unlike younger adults, do not modulate alpha power to suppress irrelevant information. Neuroimage 63 (3), 1127–1133.
URL http://dx.doi.org/10.1016/j.neuroimage.2012.07.050

Valdes-Hernandez, P. A., Ojeda-Gonzalez, A., Martinez-Montes, E., Lage-Castellanos, A., Virues-Alba, T., Valdes-Urrutia, L., Valdes-Sosa, P. A., Feb 2010. White matter architecture rather than cortical surface area correlates with the eeg alpha rhythm. Neuroimage 49 (3), 2328–2339.
URL http://dx.doi.org/10.1016/j.neuroimage.2009.10.030

van Gerven, M., Jensen, O., Apr 2009. Attention modulations of posterior alpha as a control signal for two-dimensional brain-computer interfaces. J. Neurosci. Methods 179 (1), 78–84.
URL http://dx.doi.org/10.1016/j.jneumeth.2009.01.016

Varma, S., Simon, R., 2006. Bias in error estimation when using cross-validation for model selection. BMC Bioinformatics 7 (1), 91.

Velanova, K., Lustig, C., Jacoby, L. L., Buckner, R. L., May 2007. Evidence for frontally mediated controlled processing differences in older adults. Cereb. Cortex 17 (5), 1033–1046.
URL http://dx.doi.org/10.1093/cercor/bhl013

Voelkle, M. C., Brose, A., Schmiedek, F., Lindenberger, U., May 2014. Toward a unified framework for the study of between-person and within-person structures: Building a bridge between two research paradigms. Multivariate Behavioral Research 49 (3), 193–213.
URL http://www.tandfonline.com/doi/abs/10.1080/00273171.2014.889593

Vogel, E., Machizawa, M., 2004. Neural activity predicts individual differences in visual working memory capacity. Nature 428, 748–751.

Werkle-Bergner, M., Freunberger, R., Sander, M. C., Lindenberger, U., Klimesch, W., Mar 2012. Inter-individual performance differences in younger and older adults differentially relate to amplitude modulations and phase stability of oscillations controlling working memory contents. Neuroimage 60 (1), 71–82. URL http://dx.doi.org/10.1016/j.neuroimage.2011.11.071

Worden, M. S., Foxe, J. J., Wang, N., Simpson, G. V., Mar 2000. Anticipatory biasing of visuospatial attention indexed by retinotopically specific alpha-band electroencephalography increases over occipital cortex. The Journal of Neuroscience 20 (6), RC63.

AppendixA. Spatial interpretation

AppendixA.1. Filters and patterns

The EEG surface potentials $\mathbf{x(t)} \in \mathbb{R}^U$ are believed to be a linear mixture of a set of sources $\mathbf{s}(t) \in \mathbb{R}^{\mathbf{V}}$ plus noise (e.g. Blankertz et al. (2011); Bießmann et al. (2012); Parra et al. (2005))

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) + \mathbf{n}(t).$$

U denotes the number of channels and V the number of sources. The matrix $\mathbf{A} \in \mathbb{R}^{U \times V}$ is called the forward model. Every given column $(\mathbf{a_1}, \mathbf{a_2}, \ldots, \mathbf{a_V}) = \mathbf{A}$ describes how each of the sources contributes to the surface potentials. Therefore, the column $\mathbf{a}_i$ of $\mathbf{A}$ is called the pattern of source $s_i(t)$. In the following, we will assume that for a given measurement, the time courses of all sources are stored in $\mathbf{S} = [\mathbf{s(t_1)}, \mathbf{s(t_2)}, \ldots \mathbf{S(t_T)}]$ and that the time series for all electrodes are stored in $\mathbf{X} = [\mathbf{x}(t_1), \mathbf{x}(t_2), \ldots, \mathbf{x}(t_T)]$.

If the sources and the EEG surface potentials are known, a task of interest is to find the corresponding backward model. That is, a matrix $\mathbf{W} \in \mathbb{R}^{V \times U}$ that recovers the original sources from the observed surface potentials $\mathbf{s}(t) = \mathbf{W}\mathbf{x}(t)$. An estimate $\hat{\mathbf{W}}$ can be obtained by minimizing a distance measure between the original sources $\mathbf{s}(t)$ and the reconstructed sources $\mathbf{W}\mathbf{x}(t)$. The simplest distance measure is the Euclidean distance. The resulting estimate is called least squares estimate

$$\hat{\mathbf{W}}^\top = \text{argmin}_W \sum_{t=1}^{T} (\mathbf{s}(t) - \mathbf{W}\mathbf{x}(t))^2 = \mathbf{S}\mathbf{X}^\top(\mathbf{X}\mathbf{X}^\top)^{-1}$$

The rows $(\mathbf{w_1}, \mathbf{w_2}, \ldots, \mathbf{w_s})^\top = \mathbf{W}$ of $\mathbf{W}$ are referred to as filters, as they describe the contribution of each electrode to a given source.

Typically the sources are not directly observable. Rather, a backward model $\hat{\mathbf{W}}$ that optimizes certain properties of the resulting sources is estimated. ICA (Bishop, 2007), for example, is a method to estimate a backward model that optimizes the statistical independence of the recovered sources.

After a backward model $\hat{\mathbf{W}}$ has been obtained, the least squares estimate of the corresponding forward model $\hat{\mathbf{A}}$ can be derived in analogy to the least squares estimator of the backward model and is hence,

$$\hat{\mathbf{A}} = \mathbf{X}\hat{\mathbf{S}}^\top(\hat{\mathbf{S}}\hat{\mathbf{S}}^\top)^{-\mathbf{1}} = \mathbf{X}\mathbf{X}^\top\hat{\mathbf{W}}(\hat{\mathbf{W}}\mathbf{X}\mathbf{X}^\top\hat{\mathbf{W}}^\top)^{-\mathbf{1}} \tag{A.1}$$

with $\hat{\mathbf{S}} = \hat{\mathbf{W}}\mathbf{X}$. This estimator is not invariant to constant shifts of the signals or the sources. Constant shifts in a signal originate from a constant source, which generally is not of interest in neuroscience (Parra et al., 2005). Therefore, their mean is subtracted from all rows of $\mathbf{X}$ and $\hat{\mathbf{S}}$ before the application of equation A.1. This is equivalent to including a constant source in $\mathbf{S}$ and ignoring its parameter estimate.

A particularly easy form of a backward model is a filter $\mathbf{w}^\top$, that is, a backward model that only reconstructs one source. Linear classifiers, i.e., parametric classifiers with a linear decision function of the form $f(\mathbf{x}) = \text{sign}(\mathbf{w}^\top\mathbf{x} + c)$, can be regarded as a method to obtain a filter. The optimized property of the reconstructed source is discriminability between two conditions. The filter is simply $\mathbf{w}^\top$ and the reconstructed source $\mathbf{w}^\top\mathbf{x} + c$. The corresponding pattern can be obtained by applying equation A.1.

Neuroscientific interpretations are typically facilitated by the use of patterns instead of filters (Bießmann et al., 2012; Parra et al., 2005), as patterns are not disturbed by correlated noise sources. Filters and patterns are identical (up to scaling) if $\mathbf{X}\mathbf{X}^\top$ is a multiple of the identity matrix (see equation A.1). This is equivalent to the EEG channels being uncorrelated and therefore almost never the case.

AppendixA.2. Candidate models learn a linear classifier

In this section we will show that the proposed candidate models learn a linear classifier with the entries of the between-channel covariance matrix as features. Let $\mathbf{X}(i) \in \mathbb{R}^{\mathrm{U}\times\mathrm{T}}$ be the EEG within a selected time window observed for trial i and $\hat{\mathbf{\Sigma}}(i) \in \mathbb{R}^{\mathrm{U}\times\mathrm{U}}$ the corresponding estimate of the between-channel covariance matrix. Furthermore, let $\mathbf{M}$ be the transformation matrix resulting from the composition of PCA and CSP. So, $\mathbb{R}^{2\mathrm{k}\times\mathrm{U}} \ni \mathbf{M} := \mathbf{C}\mathbf{P}$, where $\mathbf{C}$ is the transformation matrix learned by the CSP method, $\mathbf{P}$ the transformation matrix learned by PCA, and k the number of filter pairs selected for CSP. Let $\mathbf{w}'$ be the weights as learned by LLDA. Additionally, let $\text{Var} : \mathbb{R}^{\mathrm{U}\times\mathrm{T}} \to \mathbb{R}^U$ be the mapping from a matrix containing U time series of length T to the vector containing the variance of each time series. In addition to that, let $\text{diag} : \mathbb{R}^n \to \mathbb{R}^{n\times n}$ be the mapping of a vector to the corresponding diagonal matrix, $\text{diag} : \mathbb{R}^{n\times n} \to \mathbb{R}^n$ the mapping of a matrix to the vector containing the entries of its diagonal, $\text{tr}:\mathbb{R}^{n\times n} \to \mathbb{R}$ the trace of a matrix, and $\text{vec} : \mathbb{R}^{n\times p} \to \mathbb{R}^{np}$ the mapping of a matrix to a vector containing the columns of the matrix stacked on top of each other. Then the classification function $f_\theta(\mathbf{X}(i))$ for trial i is as follows. For notational clarity we drop the dependence of $\mathbf{X}$ and $\mathbf{\Sigma}$ on i.

$$
\begin{aligned}
f_\theta(\mathbf{X}) &= \text{sign}(\overbrace{\mathbf{w}'^\top \text{Var}(\mathbf{M}\mathbf{X}) + c}^{=:f'_\theta(\mathbf{X})}) && \text{(A.2)}\\
\Rightarrow f'_\theta &= \mathbf{w}'^\top \text{diag}(\mathbf{M}\hat{\mathbf{\Sigma}}\mathbf{M}^\top) + c && \text{(A.3)}\\
&= \text{tr}(\text{diag}(\mathbf{w}')\mathbf{M}\hat{\mathbf{\Sigma}}\mathbf{M}^\top) + c && \text{(A.4)}\\
&= \text{tr}(\mathbf{M}^\top \text{diag}(\mathbf{w}')\mathbf{M}\hat{\mathbf{\Sigma}}) + c && \text{(A.5)}\\
&= \underbrace{\text{vec}((\mathbf{M}^\top \overbrace{\text{diag}(\mathbf{w}')}^{\top}\mathbf{M}))^\top}_{w^\top :=} \underbrace{\text{vec}(\hat{\mathbf{\Sigma}})}_{x:=} + c && \text{(A.6)}
\end{aligned}
$$

which shows that all candidate models lead to a linear classifier with the entries of the between-channel covariance matrix as features. Hence, $\mathbf{w}$ is the filter learnt by our candidate models.

For A.3, we made use of the fact that if the covariance matrix for a random variable Y is $\mathbf{\Sigma}$, the covariance matrix for $\mathbf{M}Y$ is $\mathbf{M\Sigma M}^\top$. For A.5, we employed the fact that the trace is invariant under cyclic permutations. For A.6 $tr(\mathbf{XY}) = \mathrm{vec}(\mathbf{X}^\top)^\top \mathrm{vec}(\mathbf{Y})$ applies.

AppendixA.3. Candidate model pattern

In order to obtain the pattern corresponding to the classification function learnt by the proposed candidate models we employed equation A.1, with $\mathbf{X} = [\mathrm{Var}(\mathbf{X}(1)), \ldots, \mathrm{Var}(\mathbf{X}(I))]$ and $\hat{\mathbf{S}} = [f'_\theta(\mathbf{X}(1)), \ldots, f'_\theta(\mathbf{X}(I))]$. Prior to that, we subtracted their means from all rows of $\mathbf{X}$ and $\hat{\mathbf{S}}$ as explained in AppendixA.1. The statements in AppendixA.2 again apply. To reiterate, $\mathbf{X}(i) \in \mathbb{R}^{\mathrm{U}\times\mathrm{T}}$ is the EEG within the selected time window observed for trial i, $\mathrm{Var} : \mathbb{R}^{\mathrm{U}\times\mathrm{T}} \to \mathbb{R}^U$ is the mapping from a matrix containing U time series of length T to the vector containing the variance of each time series, and f'_θ the classification score as defined in AppendixA.2.

AppendixA.4. Proof that measurement model is not linear

Here, we show that the between-channel covariance matrix can not be expressed as a linear mapping of sources plus noise. Thus, it does not obey a linear measurement model. Let $\mathrm{Cov} : \mathbb{R}^{\mathrm{U}\times\mathrm{T}} \to \mathbb{R}^{\mathrm{U}\times\mathrm{U}}$ be the mapping from a matrix containing U time series of length T to the corresponding between-time series covariance matrix. Let $\mathbf{X}(i) \in \mathbb{R}^{\mathrm{U}\times\mathrm{T}}$ be the EEG observed for trial i, $\mathbf{S}(i) \in \mathbb{R}^{U\times T}$ the activity of the sources and $\mathbf{N}(i) \in \mathbb{R}^{U\times T}$ the noise. Dropping the dependence on i, the linear model for EEG can then be formulated as $\mathbf{X} = \mathbf{AS} + \mathbf{N}$. To simplify the proof, we assume that there is no noise. Then it follows that

$$\begin{aligned} \mathrm{Cov}(\mathbf{X}) &= \mathrm{Cov}(\mathbf{AS}) \\ &= \mathbf{A}\,\mathrm{Cov}(\mathbf{S})\mathbf{A}^\top \end{aligned}$$

which shows that the relationship between the covariances of the sources and the covariances of the electrodes is quadratic and therefore not linear. This also holds true if we drop the assumption that there is no noise.

- A method to derive person-specific models for EEG analysis is proposed.
- The approach is based on supervised learning.
- Person-specific models were more accurate than conventional nonspecific models.
- The use of individualized models for the analysis of age-differences is demonstrated.