



Toward neuroadaptive support technologies for improving digital reading: a passive BCI-based assessment of mental workload imposed by text difficulty and presentation speed during reading

Lena M. Andreassen¹  · Peter Gerjets^{2,3} · Detmar Meurers³ · Thorsten O. Zander^{1,4}

Received: 17 October 2019 / Accepted in revised form: 1 July 2020 / Published online: 1 August 2020
© The Author(s) 2020

Abstract

We investigated whether a passive brain–computer interface that was trained to distinguish low and high mental workload in the electroencephalogram (EEG) can be used to identify (1) texts of different readability difficulties and (2) texts read at different presentation speeds. For twelve subjects we calibrated a subject-dependent, but *task-independent* predictive model classifying mental workload. We then recorded EEG data from each subject, while twelve texts in blocks of three were presented to them word by word. Half of the texts were easy, and the other half were difficult texts according to classic reading formulas. From each text category three texts were read at a self-adjusted comfortable presentation speed and the other three at an increased speed. For each subject we applied the predictive model to EEG data of each word of the twelve texts. We found that the resulting predictive values for mental workload were higher for difficult texts than for easy texts. Predictive values from texts presented at an increased speed were also higher than for those presented at a normal self-adjusted speed. The results suggest that the task-independent predictive model can be used on single-subject level to build a highly predictive user model of the reader over time. Such a model could be employed in a system which continuously monitors brain activity related to mental workload and adapts to specific reader’s abilities and characteristics by adjusting the difficulty of text materials and the way it is presented to the reader in real time. A neuroadaptive system like this could foster efficient reading and text-based learning by keeping readers’ mental workload levels at an individually optimal level.

Keywords Text readability · (passive) Brain–computer interface · Mental workload · Neuroadaptive technology · User model · Online learning · Rapid serial visual presentation

1 Introduction

The emerging field of neuroadaptive technologies aims to automatically adapt a machine to changes in its operator's cognitive or affective state through an ongoing evaluation of neurophysiological activity (Krol and Zander 2017). One tool to assess information about the current mental state of a person is a passive brain–computer interface (pBCI, Zander and Kothe 2011). A pBCI is trained to distinguish between different cognitive or affective aspects of user state. Therefore, it learns to classify certain patterns of brain activity (i.e., *classes*), typically recorded through an electroencephalogram (EEG) in a controlled calibration session. An approach combining machine learning and signal filtering is used to automatically analyze a segment of EEG and to provide a guess (i.e., through interpreting the classifier output as *predictive values*) to which class the recorded activity belongs (Zander et al. 2010), resulting in a so-called *predictive model*. This model can then be applied to new data providing predictive values and hence information about the current state of the investigated cognitive process. By combining this information over time with contextual information, a predictive user model can be derived that provides meaningful details about certain aspects of higher cognition (Zander et al. 2016). This user model can then be utilized to adapt a system's behavior in a certain situation to the current state or specific aims of the user—the system becomes *neuroadaptive*. This neuroadaptation is intended to improve the interaction between user and machine by making it safer, more pleasant or more efficient (Hettinger et al. 2003; Zander 2011; Zander et al. 2016). A pBCI differs significantly from classic BCIs (i.e., active or reactive BCIs) which are built on brain activity that is generated by the operator intentionally in order to send specific control signals to a computer system, thereby replacing other means of input (e.g., keyboard) (Zander et al. 2010).

One important cognitive state that might be used for neuroadaptive technologies is *mental workload*, which is a concept that is well established in experimental psychology and human factors/ergonomics (Kramer 1991; Wickens 2008), and that is also characterized by specific neurophysiological signatures (e.g., Gevins and Smith 2000; Gevins et al. 1997; Krause et al. 2000; Pesonen et al. 2007; Scharinger et al. 2015, 2017). Recent studies claim that pBCIs can be trained to distinguish between low and high mental workload states of a user and predict mental workload in different task domains such as, e.g., n-back tasks (Grimes et al. 2008; Herff et al. 2014), span tasks (Chaouachi et al. 2011), Sternberg tasks (Baldwin and Penaranda 2012; Roy et al. 2016), arithmetic tasks (Walter et al. 2017) and visual–spatial task (e.g., mental rotation; Berka et al. 2007). However, these attempts have often been of restricted success due to several methodological limitations (cf. Gerjets et al. 2014). Specifically, efforts taken for designing a *task-independent predictive model* that can identify levels of mental workload in general, independent of the task evoking the mental activity, were not sufficiently successful. However, this is important for detecting mental workload in real-world applications, as tasks at hand can change over time. Recently, a more promising task-independent predictive model for mental workload detection has been proposed by Krol et al. (2016). This model is trained on data collected from a subject performing an arithmetic task triggering brain activity that is related to mental workload in general. While it is trained to classify these examples of mental work-

load, this model is intended to quantify mental workload in a generalized way, across tasks. This goal was not achieved by adapting feature extraction or classification algorithms in the predictive model, but by designing an experimental paradigm for the calibration session. The resulting predictive model was successfully applied to tasks in other domains, such as n-back tasks, backward span task, solving anagrams, mental rotation or understanding verbal descriptions of a complex context (Krol et al. 2016; Zhang et al. 2018). It, thus, can reliably distinguish between phases of low and high workload, not only in arithmetic tasks but also in other tasks and is therefore referred to as a task-independent predictive model. In this paper, we investigate whether this task-independent predictive model might also be suitable for a passive BCI-based assessment of mental workload imposed by text difficulty and presentation speed during reading. Workload assessment during reading in real time is a crucial prerequisite for the development of neuroadaptive support technologies that aim at improving reading efficiency in the digital age.

Reading is part of our everyday lives. Written language plays a crucial role when humans intend to share information and acquire new knowledge. From a cognitive science perspective, mental workload (in terms of working-memory load) is central for reading as reading comprehension is based on a set of processes at different levels, many of which require working-memory resources (Daneman and Carpenter 1980; Kintsch 1998; Perfetti 1985). For instance, there are reading processes at the word level (e.g., activation of word meaning), at the sentence level (e.g., integrating word meanings with the syntactical structure to a coherent sentence) and at the text level (e.g., integrating the read content plus information inferred from prior knowledge into a local and global coherent situation model). Inferences at the text level can be broadly defined as information that is not explicitly stated in the text but generated by the readers themselves from memory or the given information (Tibus et al. 2013). Inferences are crucial for understanding (“the cornerstone of comprehension”, Kendeou et al. 2016). As they are essential for comprehension, these are generated “online” during the reception process (e.g., Graesser et al. 2007). All of the above-mentioned processes—but particularly inference generation at the text level—might not only impose time demands for processing onto readers but also high levels of mental workload (depending on readers’ level of reading fluency and prior knowledge). Accordingly, the more difficult a text is and the less working-memory capacity and reading time a reader has at his disposal the more likely it is that the comprehension process might be disrupted. According to cognitive load theory, learning and comprehension will break down once the information that needs to be integrated in working memory at one point in time exceeds the available working-memory capacity (Sweller et al. 1998). In line with this assumption, reading research has shown that individual differences in working-memory capacity can account for qualitative and quantitative differences among readers in several aspects of language comprehension (Just and Carpenter 1992). In sum, the comprehension of textual information is influenced by several factors, including features of text materials itself, such as *text difficulty* (e.g., length and complexity of sentences or use of technical terms; cf. DuBay 2004; Friedman and Hoffman-Goetz 2006). Additionally, text comprehension strongly depends on characteristics of individual readers, such as their domain-specific prior knowledge (Moravcsik and Kintsch 1993), their individual working-memory capacity (Just

and Carpenter 1992), their level of reading fluency (Kuhn et al. 2010) or the time invested in reading a text (i.e., reading speed). Furthermore, these factors will interact with each other. For instance, *reading speed* will influence text comprehension performance in relation to the individual reader's reading fluency (Bell 2001; Wallot et al. 2014). Moreover, text difficulty and domain-specific prior knowledge will interact in affecting mental workload and reading comprehension. For example, the same text will impose more workload onto low prior knowledge readers than onto high prior knowledge readers (cf. McNamara et al. 1996; Sweller et al. 1998).

Therefore, developing adaptive support technologies aiming at improving reading efficiency is not trivial. This is particularly true when trying to base adaptations on measurements of individual readers' domain-specific prior knowledge, reading fluency and working-memory capacity. An alternative strategy—that we will focus on in this paper—is to base adaptive systems on an online measurement of mental workload. The reasoning for this is that mental workload during reading is directly influenced by prior knowledge, reading fluency and working-memory capacity. It is also an important factor for successful comprehension processes.

Moreover, mental workload can be used online in order to adapt the presentation of reading materials, e.g., by adjusting text difficulty. Methodologies of computational linguistics allow to automatically assess and label the complexity (i.e., *readability*) of reading materials based on linguistic features. Traditional methods use rather simple text analysis techniques that provide easily computable complexity indices (e.g., the *Flesch-Reading-Ease* (Flesch 1948) formula or the *Flesch-Kincaid-Grade-Level* formula (Kincaid et al. 1975) using easy to count text features, like the average sentence and word length). More advanced approaches to readability analysis employ a range of even more sophisticated techniques, including statistical language models (Collins-Thompson and Callan 2004; Kincaid et al. 1975; Petersen and Ostendorf 2009; Si and Callan 2001), machine learning using lexical and syntactic features (Feng 2010; Heilman et al. 2007; Schwarm and Ostendorf 2005), models integrating language-specific morphological features (Dell'Orletta et al. 2011; François and Fairon 2012; Hancke et al. 2012) and approaches emphasizing a cognitive perspective on coherence/cohesion (esp. Coh-Matrix, Graesser et al. 2014). These examples indicate that text complexity can be defined (and consequently automatically assessed) with respect to different linguistic readability metrics. These metrics range from a focus on simple linguistic entities (e.g., words) or complex, interrelated linguistic entities (e.g., sentences, paragraphs or entire documents). An automated adaptation (Zander et al. 2010) of reading materials based on mental workload could therefore be carried out by selecting text materials of an appropriate complexity level for the specific reader. This adaptation can not only be used in order to avoid reading materials that are too difficult, but also to avoid ones that are too simple, which can also impair reading comprehension (McNamara et al. 1996). An even more advanced approach would be to use algorithms for automatically quantifying text difficulty in order to correctly adjust the complexity of text materials that impose too much mental workload (Vajjala 2015; Vajjala and Meurers 2014).

Besides adapting text difficulty to mental workload, an even simpler method for the adaptation of reading materials to readers' current mental workload is to control for reading time investment by varying the speed of text presentation speed. For instance,

state-of-the-art *speed-reading applications* such as the *Spritz*TM speed-reading application (<http://www.spritzinc.com>) use *rapid serial visual presentation (RSVP)* to present text materials word for word with an adjustable presentation speed. Each word of the text is presented successively at a fixed position on the screen (Hedin and Lindgren 2007). In the Spritz application a certain letter in each word, the optimal recognition point (*ORP*, Benedetto et al. 2015; Hester et al. 2016), is highlighted in red and also presented at a fixed position to which readers should allocate their visual focus during reading to minimize eye movements, thereby reading more efficiently. Moreover, while the nominal reading speed in Spritz is set a priori by the reader, the actual reading speed differs from word to word and is influenced by factors such the length of the words and the length of the sentences, as well as by the current state of the reader. As a result, longer words are displayed longer and longer sentences have longer pauses at the end of the sentence (Benedetto et al. 2015). However, up to now, readers have to either manually adjust the nominal presentation speed of a text passage they are reading (according to their current perception of the adequacy of reading speed) or read all text passages at the same speed independent of difference in difficulty. Manual adjustment is cumbersome, requiring readers to pay attention to the current reading speed and perform manual acts to make adjustments. An adaptive system operating on the measurement of mental workload could conceivably change the nominal reading speed of text materials in a seamless way. This would facilitate an efficient reading experience that keeps readers always within an optimal range of mental workload, avoiding possible misapprehensions as well as boredom.

A pivotal prerequisite for the development of such an adaptive system is (of course) an appropriate online mental workload measure. This measure must be sensitive to variations in mental workload that are resulting from text difficulty or reading speed. Therefore, as a first step toward the development of such a system, we investigate whether a task-independent predictive model for mental workload, which is calibrated individually for any subject, can be used to distinguish between reading materials categorized as difficult or easy according to classic readability metrics. We also investigate whether different levels of reading speed induced by slow or fast text presentation can be differentiated through this predictive model.

2 Materials and methods

2.1 Subjects

Thirteen subjects (one was excluded from analyses—see Sect. 4) took part in our study, with an average age of 32.6 years ($SD = 7.4$ years). Eight of them were female. The native language of all subjects was German, and their vision was normal or corrected to normal. All subjects were naive to the experiment, and none had former experience with speed-reading applications. Prior to the experiment, subjects gave their written informed consent to participate in the study and were paid 20 euros for compensation. The experiment was conducted along a guideline ethically approved by the local ethics committee.

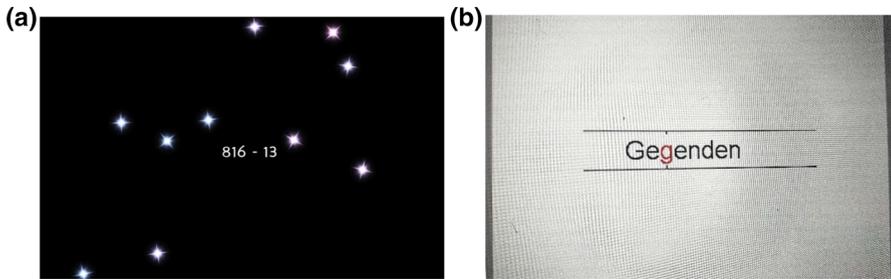


Fig. 1 Screenshots of the experimental paradigms. **a** Mental workload paradigm: high workload trial with arithmetic subtraction assignment in the center of the screen, accompanied by visual distraction represented by colorful spots randomly moving around the screen. **b** Text reading paradigm: sequentially each word is presented in the RSVP reading interface (here ‘Gegenden’ meaning ‘areas’ in German). During reading the red-highlighted letter is fixated. This letter is defined as the optimal point of recognition (ORP) for each word. The duration of presentation for each word is determined by the subjectively chosen presentation speed and factors such as word length or whether it is the end of a sentence

2.2 Apparatus

We recorded EEG from 64 active electrodes using an EEG system from Brain Products (Brain Amp DC, Brain Products GmbH, Gilching, Germany). Electrodes were positioned according to the extended international 10/20 system, with ground electrode at AFz and reference at FCz, and data were rereferenced to common average reference (CAR). Electrode impedances were kept below 5 k Ω . EEG data were recorded using the Brain Vision RDA (Brain Products GmbH, Gilching, Germany) and the LabRecorder (Delorme et al. 2010). The native sampling rate was 5000 Hz, and the recording was then subsampled to 500 Hz. The MATLAB embedded toolbox EEGLAB (Delorme and Makeig 2004) was used for EEG data analysis. All BCI classification and model application were performed using the open-source toolbox BCILAB (Kothe and Makeig 2013).

The experimental paradigm for calibrating the predictive model (*mental workload paradigm*) was implemented using SNAP (Iversen and Makeig 2014). The *text reading paradigm*, presenting the texts to the subjects, was implemented in MATLAB, using the Psychophysics Toolbox extensions (Brainard 1997; Kleiner et al. 2007) and the Spritz API. Only the reading panel of the Spritz reading interface was shown, and there was no menu to adjust reading speed or to play and pause text presentation. The ORP was highlighted in red as in the original Spritz reading interface (see Fig. 1b).

2.3 Stimuli and procedure

Subjects completed both of the above-mentioned experimental paradigms (mental workload paradigm and text reading paradigm). The mental workload paradigm (see Fig. 1a) served as calibration session for the predictive model, presenting a series of stimuli inducing low and high mental workload. This paradigm is introduced in Krol et al. (2016) and is described in more detail below. In the text reading paradigm (see Fig. 1b), subjects were asked to read twelve texts. Each text was presented at

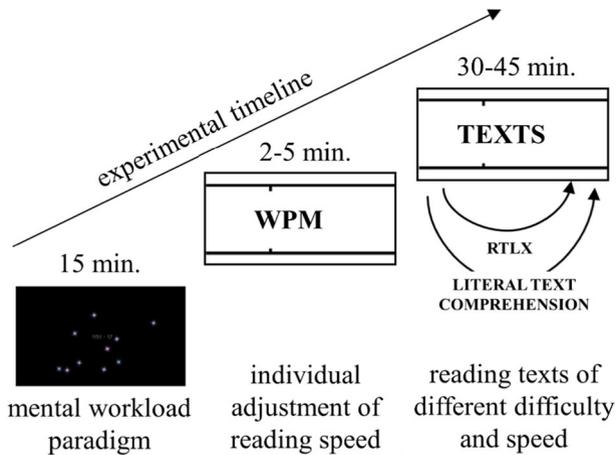


Fig. 2 Experimental procedure: Subjects first completed the mental workload paradigm. Afterward they were familiarized with the speed-reading presentation technique and adjusted the text presentation speed to a level of words per minute at which they felt comfortable reading. Then in the text reading paradigm subjects read all twelve texts in blocks of three. Block order was randomized and the three texts in a block were either easy or difficult texts presented at either the speed adjusted by the subject or with an increase of 40% in speed. After each text subjects answered three literal text comprehension questions. Following each block subjects completed the RTLX questionnaire

a certain speed (normal and fast) and had a certain difficulty (easy and difficult). For each combination of text difficulty and presentation speed three different texts were presented to the subjects, resulting in an experimental design with 3×4 texts. Details of this experimental paradigm are described below, and the timeline of the full experimental procedure is depicted in Fig. 2.

2.3.1 Mental workload paradigm

The mental workload paradigm used for calibrating the task-independent predictive model was developed by Technical University Berlin (Team PhyPA). During this paradigm subjects alternate between two tasks to induce phases of low and high mental workload. In one task subjects focused on a cross-hair in the middle of an otherwise black screen. In this condition subjects were instructed to relaxing their mind and keeping their eyes open. This task is intended to induce a low level of mental workload. To induce levels of high workload an arithmetic subtraction assignment appeared in the center of the screen. At its appearance subjects were instructed to silently subtract the number on the right side (range between 6 and 19, excluding 10 and 15) iteratively from the number on the left (range between 200 and 1200) until the assignment disappeared from the screen. In both tasks, 50% of the trials were accompanied by visual distraction represented by colorful spots randomly moving around the screen. In previous studies, these distractions have proven to serve the purpose of dissolving dependencies of any class to ocular artifacts in the EEG resulting from visual differences in the tasks (Krol et al. 2016; Zhang et al. 2018). The rationale here is that ocular artifacts are controlled by the moving distractors to a high degree. In this way both classes (low

and high workload) contain a similar, randomized amount of ocular artifacts such that the differences induced by such artifacts are not class dependent. Subjects completed 40 trials of tasks with low and 40 trials with tasks of high-induced mental workload. Each trial had a length of 10 s.

2.3.2 Text reading paradigm

Each subject read twelve texts from the GEO/GEOLino Corpus (Hancke et al. 2012) for which we calculated scores of the two common readability formulas Flesch reading ease and the Flesch-Kincaid grade level. The corpus is a collection of 1066 texts in German from the magazine GEO, which has a focus on topics from nature, culture and science and the magazine GEOLino, which covers similar topics but is targeted at young people from age 8 to 14. Therefore, texts from GEO are generally more difficult to read than those from the GEOLino magazine. Six texts were chosen from each magazine, all covering similar topics about animals and their habitat. The overall average number of words per text was 493 ($SD = 34.6$ words). GEO texts had an average word count of 472 words ($SD = 23.1$ words) and GEOLino texts of 514 words ($SD = 31.7$ words). GEO texts had an average Flesch reading ease of 45.1 ($SD = 2.4$), which is equivalent to difficult texts on college level. The Flesch-Kincaid grade level of GEO texts was 10.9 ($SD = 0.29$). For GEOLino texts the average Flesch reading ease was 62 ($SD = 1.38$), corresponding to a readability suitable for 13–15-year-old students. These texts had an average Flesch-Kincaid grade level of 7.9 ($SD = 0.24$).

All textual contents, except experimental instructions, were presented with the Spritz speed-reading application interface. In the beginning of the paradigm subjects read a short text to familiarize themselves with text reading using the Spritz interface. Thereafter, subjects read short passages of a German novel and incrementally adjusted the text presentation speed to a level they felt comfortable reading at. The initial presentation speed was 200 wpm. After each passage subjects decided by button press whether the current presentation speed appeared too slow, too fast or just right. According to the subject's decision in the first presentation, speed was altered by 100 wpm in the direction favored by the subject (or stayed the same). For the next passages, the increment was decreased by half in each step and again adjusted in the direction the subject wanted the presentation speed to alter. Presentation speed adjustment continued until the subject had decided that the current speed was comfortable for two successive text passages. The last speed value in wpm was set as the individually preferred presentation speed.

Subsequently each subject read all twelve texts. The six texts of each text difficulty category (easy vs. difficult) were randomly assigned to two groups of three texts. One of the two groups was presented in the individually adjusted presentation speed (normal category). The other group was presented at this speed raised by 40% to increase the level of mental workload (fast category). Beforehand, we assessed in a pretest ($n = 6$, age 27.2 years ($SD = 3.8$ years), all native language German, five male) that an increase in presentation speed by 40% resulted in higher subjective workload ratings according to the Raw Task Load Index questionnaire (RTLX, Hart 2006). This is a modified version of the NASA-TLX (Hart and Staveland 1988), a standardized questionnaire to assess perceived mental workload on a Likert scale along six dimensions, lacking

subjective weightings of the scales as used in the NASA-TLX. Between reading each block of three texts subjects could rest before starting the next block. The presentation order of the four text blocks was randomized. After each text block, subjects completed the RTLX questionnaire to assess their subjectively perceived workload. After each text, subjects answered three literal text comprehension questions. These questions were generated by taking facts stated within single sentences in the texts and turning them into questions (e.g., statement “With a diameter of 37 cm its eyes are the biggest in the animal kingdom.”—generated question: “Which part of the giant squid is the largest of its kind in the animal kingdom?”—displayed answers: (1) teeth, (2) nose, (3) brain. (4) eye). For each question, four possible answers were displayed only one of which was the correct choice. Subjects selected their answer by key press. In total each subject answered 36 literal text comprehension questions (3 questions \times 12 texts).

3 Data analysis

3.1 EEG data

3.1.1 Calibration of the task-independent predictive model

To calibrate a subject-dependent passive BCI classifier for the task-independent predictive model, features from the spectral domain of the individual EEG were extracted using filter bank common spatial patterns (fbCSP, Ang et al. 2008). Neuroscientific findings provide evidence that frequency bands from 4 to 7 (theta band) in fronto-central areas and 8–13 Hz (alpha band) in parietal areas of the human scalp show differences with increasing mental workload, in general (see Gerjets et al. 2014; Krause et al. 2000; Pesonen et al. 2007; Scharinger et al. 2015, 2017). Typically, increased mental workload is associated with an increase in frontal theta and a decrease in parietal alpha activity (Gevins and Smith 2000; Gevins et al. 1997).

Along these findings, each of both frequency bands was selected as a filter bank for fbCSP. In each trial, for each electrode, the EEG data were preprocessed by filtering the band power in each of the two frequency bands, serving as input for fbCSP. This approach resulted in a set of *spatial filters* to optimally extract features to discriminate between low and high workload. Each set contained 2×3 spatial filters per frequency band, where each subset of 3 filters was optimized for extracting maximal bandpower in one class while keeping the bandpower stable in the other class, reflecting the above-mentioned concept of spectral indicators for mental workload. EEG data projected through each of the 12 resulting filters served as features for classification.

Data from each trial of 10 s were cut into consecutive epochs of one second length, defining the *training trials* for classifier calibration. Each training trial served as an instance for either high or low workload, according to the class of whole trial. For each epoch, 3 features were extracted with fbCSP for each class and each frequency band, as mentioned above, resulting in a set of $(3 \times 2 \times 2 =) 12$ -dimensional feature vectors for each training trial. On the resulting set of 12-dimensional features for

each class, regularized linear discriminant analysis (LDA) was trained as a classifier that discriminates between high and low workload. The volume of the classes ranged between 1347 and 1635 samples, as each class was populated with samples from three texts and the number of words per text ranged between 449 and 545 (see second part of Sect. 2.3), according to the number of words in each text. As we expect this model to be task independent, the accuracy of the classification on the calibration data provides an estimate for the accuracy of the calibration on the text reading data. Therefore, a means of estimating classification accuracy was derived by performing a (5,5) *cross-validation* on the calibration data. The data were randomly divided into 5 partitions (indicated by the second number in (5,5)) from which each was considered to be test data, while the remaining four partitions served as calibration data for the classifier. The classifier was then trained on the calibration data and tested on the remaining partition, resulting in a classification error for this run of the cross-validation. The average of all runs of the cross-validation indicated the estimate of the classification error. This whole process was repeated 5 times (the first number in (5,5)) with different random selections of the partitions to minimize the error resulting from bias in the randomization. The average of each full cross-validation served as a final estimate of the classification error expected for the application of the classifier on the text reading data. With this, the signal processing and machine learning for the task-independent predictive model are defined.

While these filters are optimal for class discrimination, they cannot be interpreted spatially from a neuroscientific perspective. By transforming them into so-called *spatial patterns*, information about scalp projections of cortical activity that is relevant for classification can be derived and interpreted neuroscientifically (Haufe et al. 2014). For each feature a neuroscientifically interpretable pattern can be generated, identifying the electrode activity that was relevant to discriminate the classes in the according frequency band. In that way, information used by the classifier to discriminate between classes can be identified and used to validate the classification approach. Spatial patterns generated in this process were averaged over the three patterns in each subset and then over all subjects resulting in four grand average patterns displaying the topography for the two classes of low and high workload in the alpha and the theta frequency band.

3.1.2 Classifier application to text reading data

For all subjects with an estimated classification accuracy of 60% or higher the calibration process was deemed as being acceptable as it was statistically better than random (Müller-Putz et al. 2008). We used BCILAB's built in function 'onl_simulate' to apply the individually calibrated task-independent predictive model discriminating mental workload to raw data of all twelve texts per subject. The model was applied to epochs of 1 s length, serving as test trials, starting upon the onset of each displayed word of a text. The volume of the classes ranged between 1347 and 1635 samples, as each class was populated with samples from three texts and the number of words per text ranged between 449 and 545 (see second part of Sect. 2.3), according to the number of words in each text. The output of that classification was then annotated to the according

word, indicating the workload that was predicted by the model at this time. Epochs of the test trials overlapped, as there was more than one word displayed within one second during the reading task. However, deriving predictive values from overlapping epochs is unproblematic in our case. We are not performing binary classification but generating a continuous development of predictive value over time—and are not aiming at classifying single words, but rather text sections consisting of multiple words with respective workload predictive values. The resulting predictive values for each word ranged between 0 (low predicted workload) and 1 (high predicted workload). This procedure resulted in a prediction of the mental workload for each word in each text.

Subject-specific text-group averages of predictive values were compared to the intrinsic difficulty of each text by means of a two-way, repeated-measures ANOVA. The two within factors were text difficulty (easy vs. difficult) and presentation speed (normal vs. fast).

To obtain further insights into possible effects in the repeated-measures ANOVA detailed above, predictive values for each text category generated from each individual task-independent predictive model were subjected to permutation tests with 5000 permutations. This number of permutations was chosen to ensure a precision of the p value better than 0.001 in a confidence interval of 99%, according to Ojala and Garriga (2010). Predictive values from all subjects according to text difficulty (easy vs. difficult) and presentation speed (normal vs. fast) were tested against those from all other categories, resulting in $6 \left(= \binom{4}{2} = 3 \times 2 \times 1 \right)$ permutation tests. Five of these tests were one-tailed because we assumed that easy texts would result in lower predictive values than difficult texts. Also, predictive values from texts of comparable difficulty were expected to be lower when presented at normal speed than predictive values of texts presented at an increased speed. Easy texts presented at normal speed were expected to have lower predictive values than those of difficult, fast presented texts. As no assumption regarding mental workload could be made, the permutation test comparing predictive values of easy/fast speed texts and difficult/normal speed texts was two-tailed.

Finally, we analyzed the predictive values on single-subject level by performing further one-tailed permutation tests. We hypothesized that predictive values for easy texts would be lower than those of difficult texts when read at the same presentation speed. Further we expected texts of the same text difficulty to have lower workload predictive values for texts read at normal speed than at a faster pace. Finally, we expected easy texts read at normal speed to have lower predictive values than difficult texts read at fast speed. The five resulting permutation tests per subject were computed with again 5000 permutations using all predictive values from the three texts belonging to the same text category.

To obtain further insight into classification performance on a single-subject level we performed an analysis of how to determine the amount of read words that need to be classified to make a reliable discrimination between text categories. For each word position we averaged the predictive values derived from texts of the same category (three for each text difficulty or speed) at this position to correct for the intra-text

variance of word and text complexity at that position in that category. This allows us to calculate the *cumulative predictive value* for each word position in each category by averaging these corrected predictive values up to the selected word position: For each word w_n^k in text T_k , that is from a certain text category c , the cumulative predictive value cp at position n is derived by

$$cp_n^c = \sum_{l=1}^n \sum_{k=1}^3 \frac{cla(w_l^k)}{3n}$$

where $cla(w_l^k)$ denotes the predictive value derived for the word w_l^k by linear discriminant analysis. Let E be the segment of EEG data recorded for word w_l^k and P be the preprocessing of that data as described in Sect. 3. Then the application of filter bank CSP on the preprocessed data results in a 12-dimensional feature vector $fbCSP(P(E))$, as described in Sect. 3. Applying LDA to this feature vector then results in the single-digit predictive value $cla(w_l^k) = LDA(fbCSP(P(E)))$.

We plotted the development of the cumulative predictive values $(cp_n^c)_{n \in \{1, \dots, 422\}}$ for each word position in each text category. As the shortest text contained 422 words we only considered the cumulative values of each text category up to the 422nd word. We plotted all four combinations of factor comparisons (difficulty (easy vs. difficult) vs. presentation speed (normal vs. fast)) for each subject. Each trained task-independent predictive model is a probabilistic operator with a certain error rate. Hence, each (cp_n^c) should only be considered as an estimate of the actual workload at that word position. We calculated the individual variance of the error of the predictive values for each subject by applying the mental workload classifier again to the calibration paradigm data. The standard deviations of the predictive values for each class were averaged and served as a criteria for the error range for the cumulative predictive values $(cp_n^c)_{n \in \{1, \dots, 422\}}$: For each subject the *classifier deviation* is defined as $D := \frac{\text{std}(C^1) + \text{std}(C^2)}{2}$, where $\text{std}(C^k)$ denotes the standard deviation of the predictive values for class k . Since it is taking all predictive values considered into account, cp_{422}^c can be seen as the most informed estimate of the workload associated with that text category. The *stability word position* W is defined as the highest word position where the cumulative predictive value exceeds the (rather conservative) criteria of one quarter of the classifier deviation from cp_{422}^c : For each subject and text category c , $W^c := cp_M^c$, with $M = \max_n (|cp_n^c - cp_{422}^c| > \frac{D}{4})$. The cumulative predictive values for each word position following W^c can then be considered to be an estimate for the workload associated to the corresponding text that is as good as that from cp_{422}^c . In other words, the text position of the stability word serves as an estimate of how many words of a text a single person must read such that the task-independent predictive model can provide a best estimate for mental workload induced by that particular text category. For each subject, we plotted the highest text position from the four determined stability word positions (for the four different text categories) as the general stability word position. We assume that after this text position, we can use the cumulative value at this text position as a reliable indicator of induced mental workload for all text categories.

3.2 RTLX

Subjective workload ratings assessed in the RTLX questionnaire were converted to workload scores according to NASA-TLX procedures (Hart 2006). These scores were then subjected to a two-way repeated-measures ANOVA. The two within factors are text difficulty (easy vs. difficult) and presentation speed (normal vs. fast).

3.3 Literal text comprehension

For each subject we collected the number of incorrect answers (total of 9 questions per block, three questions per text) per text category. We then computed a two-way repeated-measures ANOVA, with the two within factors text difficulty (easy vs. difficult) and presentation speed (normal vs. fast) for the counts of correctly answered questions.

4 Results

4.1 EEG data

4.1.1 Mental workload paradigm

Table 1 gives an overview of classification error rates estimated from cross-validation in the mental workload paradigm of all initial thirteen subjects. The average estimated classification error rate was 25.9% (SD = 4.9). We performed a check to see whether individual results provide evidence that the task-independent predictive model had classification accuracies that are significantly better than random classification, taking the confidence interval into account (see Müller-Putz et al. 2008 for a detailed description of this approach). The results show that, for all but one subject, classifications are expected to be clearly better than random classification. Therefore, this subject was excluded from further analysis.

Figure 3 depicts that the spatial patterns resulting from the classifier model show a clear focus on fronto-central electrodes for the theta band and on parietal electrodes for the alpha band, providing evidence that the filters derived by fbCSP are in accordance with the underlying neurophysiology.

4.1.2 Text reading paradigm

Subject-specific text-group averages of predictive values are displayed in Fig. 4.

The two-way repeated-measures ANOVA (see Fig. 6a) performed on mental workload predictive values yielded significance for the main factor text difficulty, $F(1,11) = 5.04$, $p = 0.046$. Predictive values of difficult texts ($M = 0.618$, $SD = 0.272$) were higher than predictive values for easy texts ($M = 0.553$, $SD = 0.259$). The main factor text presentation speed was significant as well, $F(1,11) = 11.64$, $p = 0.006$. Predictive values of texts presented at normal speed were lower ($M = 0.559$, $SD = 0.267$)

Table 1 Estimated classification results of the mental workload paradigm for all initial thirteen subjects

Subject	ER (SD)
1	22.3 (2.0)
2	31.0 (12.0)
3	13.9 (3.1)
4	25.0 (3.4)
5	27.1 (3.2)
6	28.6 (6.5)
7	10.9 (5.5)
8	27.8 (3.2)
9	31.8 (2.4)
10	27.0 (8.4)
11	35.8 (5.7)
12	13.3 (4.2)
13	41.8 (4.1)
Average	25.9 (4.9)

Obtained error rates (ER) from (5,5) cross-validation in percent and their standard deviations (SD), resulting from variance in the classification error estimated from each run of the cross-validation, are reported

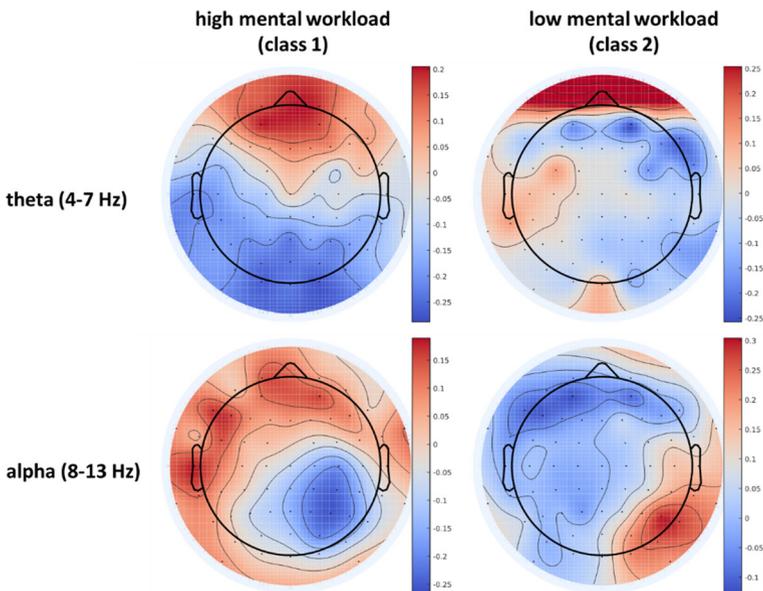


Fig. 3 Spatial patterns available to the mental workload task-independent predictive model averaged over all twelve subjects where classification was discriminative. For the high mental workload condition (class 1) the patterns show a clear increase in frontal theta and a decrease in (lateral) parietal alpha. These results are in accordance with findings from scientific literature regarding the effects of increased mental workload

than those of texts presented at the increased speed ($M = 0.616$, $SD = 0.264$). There was no significant interaction, $F(1,11) = 0.036$, $p = 0.854$.

The results of five from the six permutation tests across subjects were highly significant (all $ps < 0.0001$). Only in the case where predictive values of easy texts presented

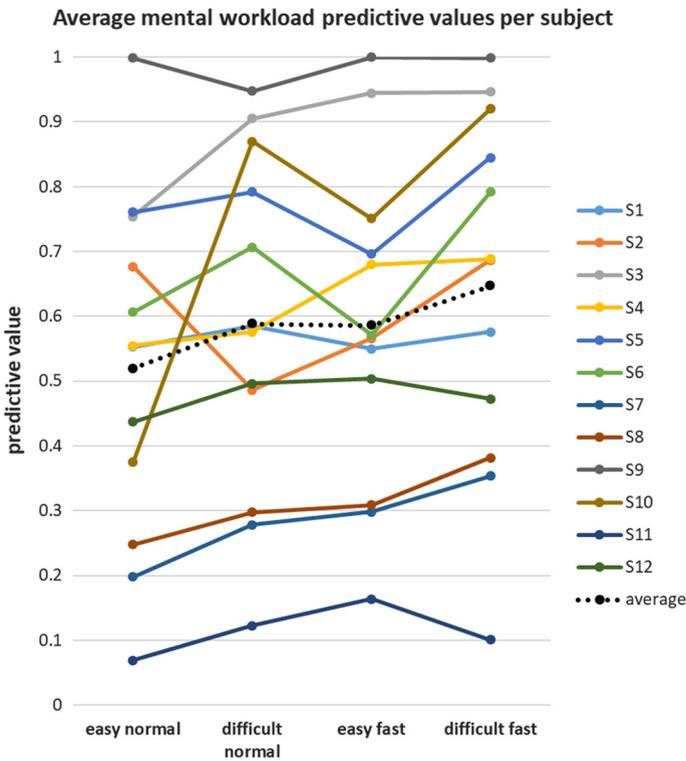


Fig. 4 Subject-wise predictive values of mental workload derived from application of the task-independent predictive model trained on data from the mental workload paradigm to text reading data. For each text reading condition all predictive values of all three texts belonging to the respective condition were averaged. The dotted line displays the grand average of all subjects ($n = 12$). The results of the ANOVA and the related permutation tests across subjects are validating the general hypotheses (as depicted in the dotted line), while accordingly significant differences between conditions within subjects (solid lines) were also found in subject-specific permutation tests

at fast speed tested against predictive values from difficult texts presented at normal speed, results were not significant ($p = 0.340$). However, absolute values of observed differences in predictive values between text conditions ($M = 0.058, SD = 0.036$) were much smaller in all tests than variances of predictive values within text conditions ($M = 0.265, SD = 0.011$). Effect sizes were accordingly small ($M = 0.181, SD = 0.069$), ranging between 0.129 (easy fast vs. difficult fast) and 0.301 (easy normal vs. difficult fast) in tests with significant results.

On single-subject level, one-tailed permutation tests (see Table 2) showed significant results for 9 of 12 subjects when predictive values of easy texts and difficult texts, both read in normal speed, were tested against each other. Nine tests also yielded significance when easy texts read at normal speed and difficult texts read fast were compared. When we compared predictive values from difficult texts read at normal and fast speed, 9 of 12 permutation tests were significant. For easy texts read in normal and fast speed 8 of 12 tests yielded significance. The lowest count of significant

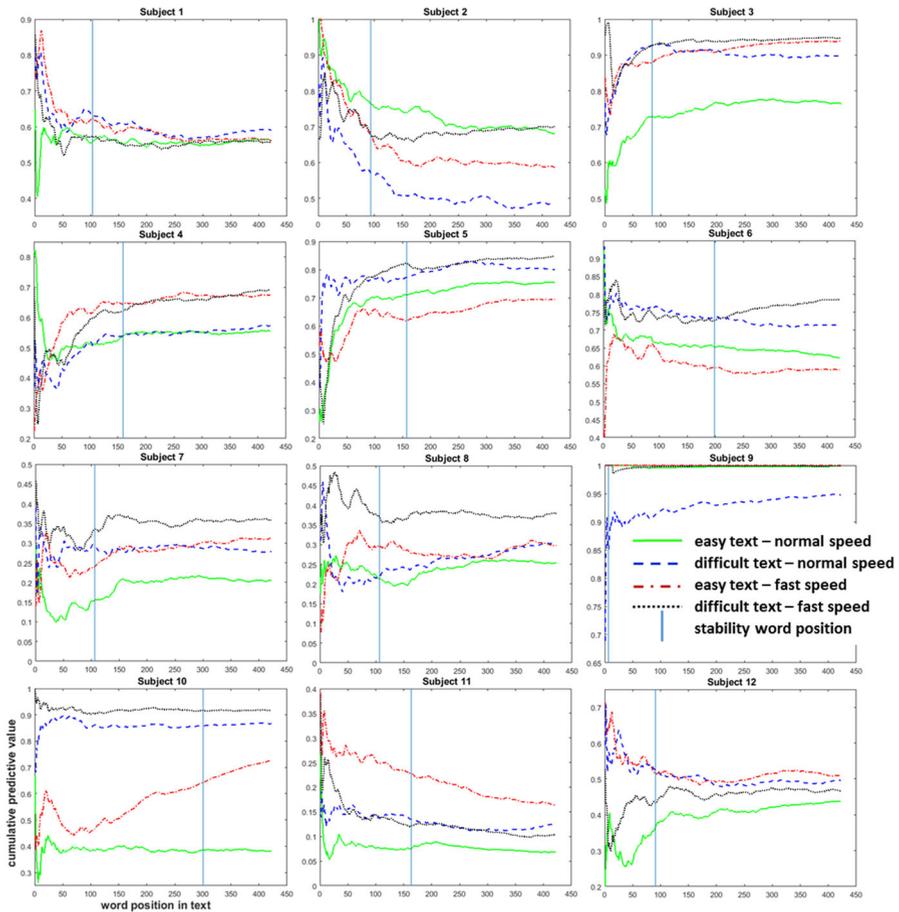


Fig. 5 Cumulative predictive values derived from the predictive models of the mental workload classifiers. Predictive values of all three texts per category were averaged and cumulated for each successive word in the texts. The development of cumulative predictive values for all four text reading conditions is shown for the first 422 words in the texts. On the x-axis the text position of the word is displayed. On the y-axis the averaged cumulative predictive values are noted. The individual stability word position (highest of all for text categories) is indicated by a vertical black line

permutation tests was reached when we tested predictive values of easy and difficult texts read at both fast speed against each other—here only 7 of 12 tests had significant results. The average number of significant permutation test results was 3.67 (SD = 1.07) per subject out of 5 tests. Overall effect sizes of significant permutation test results are close to medium with an average effect size of -0.444 (SD = 0.379). The individually determined stability word position (see Fig. 5) was at text position 130.83 (SD = 72.72) on average.

Table 2 Results of subject-wise one-tailed permutation tests performed on predictive values

VP	Permutation test result	Easy/normal versus difficult/normal	Easy/fast versus difficult/fast	Difficult/normal versus difficult/fast	Easy/normal versus easy/fast	Easy/normal versus difficult/fast	Stability word position	Workload paradigm classification error rate (SD) in percent
1	<i>p</i> value	<0.005	<0.01	>0.01	>0.01	>0.01	103	22.3 (2.0)
	Observed difference	-0.0317	-0.0263	0.0086	0.0032	-0.0230		
2	Effect size	0.1040	0.0892	0.0289	0.0108	0.0778		
	<i>p</i> value	>0.01	<0.0005	<0.0005	>0.01	>0.01	94	31.0 (12)
3	Observed difference	0.1904	-0.1209	-0.2004	0.1109	-0.0010		
	Effect size	0.5113	0.3345	0.5445	0.3032	0.0285		
4	<i>p</i> value	<0.0005	>0.01	<0.0005	<0.0005	<0.0005	85	13.9 (3.1)
	Observed difference	-0.1513	-0.0019	-0.0407	-0.1901	-0.1920		
5	Effect size	0.7782	0.0163	0.2854	1.1114	1.1178		
	<i>p</i> value	>0.01	>0.01	<0.0005	<0.0005	<0.0005	159	25 (3.4)
6	Observed difference	-0.0216	-0.0086	-0.1122	-0.1253	-0.1338		
	Effect size	0.0666	0.0281	0.3556	0.3989	0.4218		
7	<i>p</i> value	<0.005	<0.0005	<0.0005	>0.01	<0.0005	157	27.1 (3.2)
	Observed difference	-0.0308	-0.1484	-0.0530	0.0645	-0.0839		

Table 2 continued

VP	Permutation test result	Easy/normal versus difficult/normal	Easy/fast versus difficult/fast	Difficult/normal versus difficult/fast	Easy/normal versus easy/fast	Easy/normal versus difficult/fast	Stability word position	Workload paradigm classification error rate (SD) in percent
6	Effect size	0.1059	0.5330	0.1100	0.2121	0.3103		
	<i>p</i> value	<0.0005	<0.0005	<0.0005	>0.01	<0.0005	199	28.6 (6.5)
7	Observed difference	-0.1005	-0.2209	-0.0851	0.0353	-0.1856		
	Effect size	0.3661	0.8609	0.3438	0.1245	0.7081		
8	<i>p</i> value	<0.0005	<0.0005	<0.0005	<0.0005	<0.0005	106	10.9 (5.5)
	Observed difference	-0.0805	-0.0553	-0.0752	-0.1003	-0.1557		
9	Effect size	0.3018	0.1902	0.2655	0.3654	0.5733		
	<i>p</i> value	<0.0005	<0.0005	<0.0005	<0.0005	<0.0005	106	27.8 (3.2)
10	Observed difference	-0.0498	-0.0730	-0.0839	-0.0607	-0.1337		
	Effect size	0.1554	0.2152	0.2502	0.1870	0.4068		
10	<i>p</i> value	>0.01	>0.01	<0.0005	<0.001	>0.01	7	31.8 (2.4)
	Observed difference	0.0512	0.0013	-0.0512	-0.0013	0.0001		
10	Effect size	0.5696	0.0969	0.5471	0.1264	0.0023		
	<i>p</i> value	<0.0005	<0.0005	<0.0005	<0.0005	<0.0005	300	27.0 (8.4)
10	Observed difference	-0.4952	-0.1691	-0.0505	-0.3765	-0.5457		

Table 2 continued

VP	Permutation test result	Easy/normal versus difficult/normal	Easy/fast versus difficult/fast	Difficult/normal versus difficult/fast	Easy/normal versus easy/fast	Easy/normal versus difficult/fast	Stability word position	Workload paradigm classification error rate (SD) in percent
11	Effect size	1.4862	0.5934	0.2212	0.9657	1.7869		
	p value	<0.0005	>0.01	>0.01	<0.0005	<0.0005	163	35.8 (5.7)
	Observed difference	- 0.0536	0.0632	0.0220	- 0.0948	- 0.0316		
12	Effect size	0.3161	0.3301	0.1217	0.5256	0.1979		
	p value	<0.0005	>0.01	>0.01	<0.0005	<0.005	91	13.3 (4.2)
	Observed difference	- 0.0592	0.0314	0.0241	- 0.0665	- 0.0351		
	Effect size	0.1845	0.0963	0.0730	0.2099	0.1071		
		9 of 12	7 of 12	9 of 12	8 of 12	9 of 12		

For each test the category first named in the table was expected to have lower workload predictive values. We also noted the individual stability word position for predictive values (indicating how many words need to be read to make a reliable estimation of the average predictive value the reading of a text induces in the specific subject), as well as the classification error rate of the mental workload classifiers. The tests were Bonferroni corrected for multiple testings in each subject, resulting in a significance level of 0.01

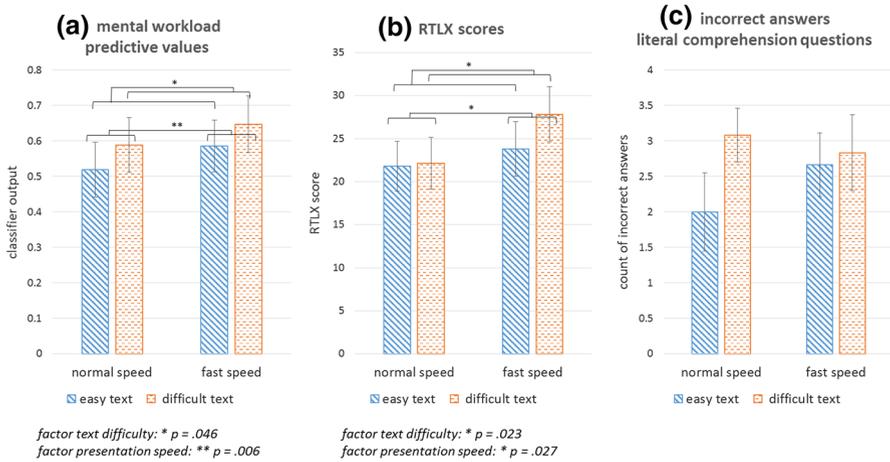


Fig. 6 **a** Predictive value averages ($n = 12$) of the task-independent predictive model applied to EEG data of each word of the twelve texts subjects read, **b** averages ($n = 12$) of RTLX ratings for the different text difficulties and text presentation speeds and **c** averaged ($n = 12$) numbers of incorrectly answered literal text comprehension questions. Error bars represent standard deviations. Significant results of the ANOVAs are indicated

4.2 RTLX

The two-way repeated-measures ANOVA (see Fig. 6b) performed on subjective workload ratings revealed significance for the main factor text difficulty, $F(1, 11) = 6.98$, $p = 0.023$. Overall difficult texts ($M = 24.98$, $SD = 10.76$) received higher mental workload ratings than easy texts ($M = 22.78$, $SD = 10.51$). Results for the main factor presentation speed were significant as well, $F(1, 11) = 6.48$, $p = 0.027$. Texts presented at the normal ($M = 21.95$, $SD = 10.18$) self-adjusted reading speed received lower RTLX ratings than texts presented with an increase of 40% ($M = 25.81$, $SD = 11.09$). The interaction was not significant, $F(1, 11) = 1.87$, $p = 0.199$.

4.3 Literal text comprehension

Individually adjusted text presentation rates from all subjects ranged between 150 and 500 wpm. The average adjusted reading speed was 269 wpm ($SD = 109$ wpm). The ANOVA performed on incorrect answers given to literal text comprehension questions (see Fig. 6c) revealed neither significant main effects, nor an interaction effect, all p s > 0.249 . On average subjects answered 2.6 ($SD = 1.7$) questions out of nine per text block incorrectly. An average of 2 ($SD = 1.8$) incorrect answers was given for easy texts and 3.1 ($SD = 1.3$) for difficult texts presented at normal speed. For texts blocks with an increased presentation speed, questions on easy texts were answered 2.7 ($SD = 1.5$) times incorrectly on average and difficult texts 2.8 ($SD = 1.9$) times.

5 Discussion

The aim of this investigation was to examine whether a task-independent predictive model for mental workload detection shows a sensitivity high enough to identify the individual difficulty induced by reading a specific text. The results of the presented experiment show that an individually calibrated model shows significant differences between reading easy and difficult texts at different presentation speeds.

Regarding text readability, results from EEG data suggest that we can significantly distinguish easy from difficult texts, even on single-subject level, even though variances of mental workload predictive values within and between subjects are high. High variances between subjects most probably are a result of different intra-subjective ratios between arithmetic and text reading abilities.

To our knowledge no other study has yet investigated whether text difficulty and reading speed can be directly deduced from mental workload assessed by EEG. Studies (e.g., Schultheis and Jameson 2004) have investigated changes in EEG activity induced by reading. They provided a descriptive analysis of that brain activity, but did neither identify sensitive and specific cognitive processes—like mental workload—related to that activity, nor did they define and investigate a measure from which text difficulty or reading speed could be deduced.

The neuroadaptive measure defined here showed overall that predictive values were higher for words from difficult texts than for words from easy texts. Moreover, on single-subject level for nine out of twelve subjects, mental workload predictive values were significantly higher for difficult texts than for easy texts when reading at a self-adjusted speed level. If difficult texts were read with an increased speed, predictive values for mental workload were even higher and more distinct from the lower predictive values of easy texts read at normal speed. The latter was the case for nine out of twelve subjects as well. When we compared predictive values of easy and difficult texts both read at an increased speed, only seven out of twelve permutation tests yielded significant results.

Regarding differences in mental workload induced by different text presentation speeds we found a highly significant effect for increased presentation speeds on mental workload predictive values. On single-subject level, in nine out of twelve cases mental workload predictive values were higher for difficult texts when they were read at an increased speed compared to reading at a self-adjusted speed level. For easy texts, in eight out of twelve subjects' predictive values were considerably higher when reading at increased texts presentation speed opposed to reading at a self-adjusted speed.

As the chosen texts were written for specific target groups (educated adults and young teenagers), it can be assumed that the difficulty of each category is aimed at a certain educational level. Furthermore, measures for text complexity confirmed a difference between easy and difficult texts (see Sects. 2, 2.3, 2.3.2) and an increase of reading speed should add a stable amount of mental workload for the reader. Taken together, it can be assumed that each full text induces a certain level of mental workload when read, and that these levels can be identified by the used predictive models. The classifier outputs are in line with this assumption as shown in Fig. 5 and Table 2. The curves in Fig. 5 stabilize after about 130 words, showing that the classifier outputs converge to a clear estimate of the induced mental workload, already after about 30 s

of reading. In addition, a clear, expected difference in measured workload can be seen across subjects between the different text categories, as validated by the ANOVA and depicted in Fig. 4. In conclusion, the results of this study are fully in line with the expected effect of the stimuli on the mental workload of the readers, providing a high validity to the used task-independent approach based on passive BCI.

Results of the ratings given in the RTLX questionnaire assert that readers perceive a rise in reading speed of 40% as an increase in subjectively experienced mental workload. Also, reading difficult texts was associated with higher scores for mental workload as compared to easier texts. We therefore conclude that the subjective experience of mental workload was in accordance with the properties of the predictive values we obtained from the predictive model for each subject with regard to changes in text readability and reading speed.

Since no statistical differences in literal comprehension were shown between different text difficulties and presentation speeds, it can be concluded that an increased presentation speed did not lead to lower levels of literal comprehension in our study. On average, two-thirds of the comprehension questions within each text block were answered correctly by subjects. Thus, the results from literal comprehension questions indicate that subjects read all variations of texts attentively and with the same level of basic understanding. However, as we did not include more complex comprehension questions that require for elaborated inferences, we cannot rule out that higher comprehension processes might have suffered from increased reading speeds.

The self-selected presentation speed for the normal presentation showed a large variance between subjects. The aim of selecting the individual speed was to generate a speed that is comfortable for each individual reader, neither overloading nor boring them. The results of the text comprehension analysis showed a high comprehension rate with low variance, indicating that subjects could follow the texts easily. This leads to the conclusion that the presentation speeds were indeed comfortable for the subjects, even though the presentation speeds differed strongly between readers.

The overall duration of our experiment was a little less than an hour on average. We took several counter measures to avoid fatigue overall, as well as possible effects of it. Firstly, participants were allowed several breaks throughout the experiment to recover from tasks they were performing. Even if fatigue or high workload was not avoided by those breaks, they should not have an effect on the results we saw from the experiment, as the order the four text categories presented to subjects was randomized and the overall duration of the experiment was about the same for all subjects, so the induced fatigue should not vary among subjects.

Altogether, the results showed that the task-independent predictive models are capable of identifying different levels of text readability and presentation speed on single-subject level. We proposed an estimate of how many words need to be read to make a reliable estimation of the average mental workload associated with a text. This criterion suggests that an average count of about 130 words would be sufficient.

In conclusion, the presented results allow to add text readability and presentation speed to the task domains the task-independent mental workload classifier introduced in Krol et al. (2016) and further investigated by Zhang et al. (2018) can be successfully applied to.

The presented approach deviates from often-used practices of binary classification based on BCI and aims at the identification of the levels of workload on a continuous scale. Results presented here support the idea that, in future applications of this approach, the subjective reading difficulty of a certain section in a text can be identified based on estimations of the induced mental workload. It, hence, is not necessary to define certain categories of difficulty for texts, as done in the presented experiment. Instead, the subjective, momentary reading difficulty of a text can be traced over time on a continuous scale.

However, it remains unclear how well the predictive model would perform for texts whose readability is more similar than that of the texts used in this study. The precision of the mental workload classifier should be further investigated in future research by presenting subjects with a larger spectrum of readability levels.

Furthermore, smaller steps of presentation speed alternation should be examined in order to investigate how mental workload predictive values are changed at presentation speeds lower than an individually adjusted level. It would also be of interest to determine whether or not predictive values behave linearly with changes in text readability and presentation speed.

In a real-world reading scenario, the main goal of the user is to read a specific text. From a user experience point of view, it would be bothersome to have to invest additional time and effort to enable a more individualized user support for the main task of reading. For the implementation of a neuroadaptive reading interface currently such additional effort must be invested by the user, as he undergoes an initial calibration phase to collect training data for the task-independent mental workload classifier. In the future, this time-consuming step in creating a neuroadaptive reading environment could become unnecessary through the establishment of a mental workload classifier that is not only task-, but subject-independent as well. In such a universal classification (Zander 2011, p. 96ff, p. 142ff.) a classifier can be defined that can be transferred between sessions and even between subjects, leading to reduced calibration time for an individual subject or even making it obsolete. Approaches toward such universal classification focusing specifically on mental workload are currently investigated (see e.g., Durkee et al. 2016; Zander 2011). Another approach for reduction of calibration time is transfer learning between calibration sessions (Jayaram et al. 2016) which has already been applied to motor imagery classification contexts (Krauledat 2008, p. 90ff). In future investigations such approaches for a reduction of calibration time should be investigated for the here presented neuroadaptive reading environment.

In addition, it would be of interest how well the results from this study can be generalized to other forms of reading and text presentation techniques (as e.g., traditional continuous text reading). We assume that in further studies other reading scenarios could be investigated by additional collection of eye-tracking data. This would enable us to map which word of a text currently is read by a subject to the corresponding EEG data we record.

Though exceeding the scope of this study, it should be further investigated in what way mental workload is effected by RSVP reading itself. It has already been found that workload can be increased when long text passages are read with RSVP presentation (Öquist and Goldstein 2003) and mental workload in general is higher for RSVP reading compared to traditional (continuous) reading (Gannon et al. 2016). An analysis

and comparison of RSVP reading and traditional reading regarding mental workload is not a part of our work presented here as it is outside of the realm of the research questions we pose, which is restricted to RSVP reading only. In a follow-up study, where the generalizability of our results to traditional reading will be tested, this comparison of reading techniques would be investigated in more depth.

Moreover, in this feasibility study the subject's reading fluency was not assessed. Hence, possible effects of reading fluency on the classifier performance and workload prediction were not analyzed. In a follow-up study with more subjects the effects of reading fluency should be investigated.

While effect sizes on individual data are at a good level, those between subjects are rather small. This can be seen as a result of the high variance of the predictive values between subjects, as shown in Fig. 4. Currently, the reason for this high variance is unclear. We assume, though, that it is a result of differences in mathematical abilities and reading skills, which should be investigated in future studies.

As reproducibility of results is relevant, specifically for new approaches as presented here, all preprocessed data and analyses code will be linked at the website of the Society for Neuroadaptive Technology (www.neuroadaptive.org).

6 Conclusion

Neuroadaptive systems have the potential, if well designed, to decrease mental workload for users during interaction with a system (Byrne and Parasuraman 1996). Such a system can continuously learn, while the user interacts with the machine and thereby build and refine a user model that allows for individualization of software and contents over time.

Such a model would reflect highly subjective aspects of information that are not typically included in methods relying on large amounts of objective data. Statistical inference (see Vapnik 1999) based on large amounts of data can be used to calculate general readability metrics, but leave no room for adaptations based on individual, momentary situational changes (such as changes in a person's mindset). It is precisely this kind of transient information, however, that is paramount for correct situational interpretation in general (Hudlicka and Mcneese 2002). We propose that the approach presented here could be used to complement models built through statistical inference on large amounts of data by providing information about the subjective, momentary interpretation of the human. Technology could then not only learn to understand general patterns that exist in our world, but also gain a better understanding of the momentary mindset of individual human beings in this world. The study presented here provides an additional, supportive argument—with the example of mental workload—for this vision.

We demonstrate that predictions made by the task-independent predictive model trained to identify levels of workload can be utilized in order to assess differences in text readability on a single-subject level. Additionally, changes in mental workload induced by different levels of text presentation speeds are also reflected in the output of the model. These levels can be distinguished through the classifier's predictive values of about the first 130 words read, which took about 30 s in our experiments. With

that, the system can gain an understanding of the current effect a text has toward the reader's cognition and adapt itself accordingly.

Even though questions regarding the precision of the mental workload classifier in practical contexts remain to be answered, the results of our study suggest possible benefits this measure could bring in the form of a neuroadaptive system for (speed) reading applications.

A user model with a classifier trained over time to identify text readability levels for an individual user could automatically adapt parameters of the reading application in real time, e.g., the presentation speed. Conducting such adaptations within a closed-loop system between reading application and reader can elicit changes in the user's mental state. These changes can then be fed back to the system again to perform further adaptation or to learn from the induced changes in mental user state for future adaptation. The implementation of a neuroadaptive reading application could enable more effective, individualized reading at a balanced level of workload. Readers should neither be over- nor underloaded while reading for an optimal state of mind to process and store the information read. This was firstly discussed in instructional design theories (e.g., Vygotsky 1980 or Salomon 1984). The main problem for the application of these theories is to measure the current, actual mental workload, which was difficult so far. The idea of using neuroadaptive interfaces to solve this problem is discussed for the first time in a general sense in Gerjets et al. (2014). The work presented here builds up on these concepts and describes an approach that allows for the assessment of mental workload while reading a text. The next steps will be the use of this information in an adaptive reading interface, closing the loop and keeping the reader in an improved cognitive state, which can be done in different ways. Apart from presentation speed changes, possible adaptation of readability levels might include real-time exchange of words or restructuration of the text material. If the system has, e.g., detected that the current material is difficult for the user to read, infrequent words could be automatically exchanged by easier synonyms. Furthermore, sentences could be restructured in order to reduce sentence length. Further possibilities for real-time text simplification have been discussed in Crossley et al. (2011).

This study shows a new example of how passive brain-computer interfaces can be used to develop innovative neuroadaptive technologies. It also suggests ways how such technology could be implemented, inspiring studies that are not aiming at evaluating BCIs, but furthermore investigating how this approach can be integrated into technologies to support people in their daily life. Neuroadaptive systems capable of inducing user-specific adaptation to the system's behavior might not only improve reading experience but could be particularly useful in any workload-sensitive context, such as personalized learning environments or online tutoring systems, leading to new ways for technology to present information to us.

Acknowledgements Open Access funding provided by Projekt DEAL. We especially thank Spritz™ for providing us with the API of their speed-reading application to be integrated in our experimental reading paradigm. Moreover, we thank the authors of Krol et al. (2016) for making the experimental paradigm for calibration of the task-independent predictive model available to us. We thank Dr. David E. Medine for proofreading and general suggestions for improving the manuscript. Part of this work was supported by the Deutsche Forschungsgemeinschaft (ZA 821/3-1). This research was also supported by the LEAD Graduate School (GSC1028), a project of the Excellence Initiative of the German federal and state governments. TOZ

was supported by the Center for Bioelectric Interfaces NRU Higher School of Economics, RF Government grant, AG. No. 14.641.31.0003.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Ang, K.K., Chin, Z.Y., Zhang, H., Guan, C.: Filter bank common spatial pattern (FBCSP) in brain–computer interface. In: IEEE International Joint Conference on Neural Networks. IJCNN 2008 (IEEE World Congress on Computational Intelligence), pp. 2390–2397. IEEE (2008)
- Baldwin, C.L., Penaranda, B.N.: Adaptive training using an artificial neural network and EEG metrics for within- and cross-task workload classification. *Neuroimage* **59**(1), 48–56 (2012)
- Bell, T.: Extensive reading: speed and comprehension. *Read. Matrix* **1**(1) (2001)
- Benedetto, S., Carbone, A., Pedrotti, M., Le Freve, K., Bey, L.A.Y., Baccino, T.: Rapid serial visual presentation in reading: the case of Spritz. *Comput Hum Behavior* **45**, 352–358 (2015)
- Berka, C., Levendowski, D.J., Lumicao, M.N., Yau, A., Davis, G., Zivkovic, V.T., Olmstead, R.E., Tremoulet, P.D., Craven, P.L.: EEG correlates of task engagement and mental workload in vigilance, learning, and memory tasks. *Aviat. Space Environ. Med.* **78**(5), B231–B244 (2007)
- Brainard, D.H.: The psychophysics toolbox. *Spat. Vis.* **10**, 433–436 (1997)
- Byrne, E.A., Parasuraman, R.: Psychophysiology and adaptive automation. *Biol. Psychol.* **42**(3), 249–268 (1996)
- Chouachi, M., Jraidi, I., Frasson, C.: Modeling mental workload using EEG features for intelligent systems. In: International Conference on User Modeling, Adaptation, and Personalization, pp. 50–61. Springer, Berlin (2011)
- Collins-Thompson, K., Callan, J.: A language modeling approach to predicting reading difficulty. In: Proceedings of HLT/NAACL 2004, Boston, USA (2004)
- Crossley, S.A., Allen, D.B., McNamara, D.S.: Text readability and intuitive simplification: a comparison of readability formulas. *Read. Foreign Lang.* **23**(1), 86 (2011)
- Daneman, M., Carpenter, P.A.: Individual differences in working memory and reading. *J. Verbal Learn. Verbal Behav.* **19**(4), 450–466 (1980). [https://doi.org/10.1016/S0022-5371\(80\)90312-6](https://doi.org/10.1016/S0022-5371(80)90312-6)
- Dell'Orletta, F., Montemagni, S., Venturi, G.: READ-IT: assessing readability of Italian texts with a view to text simplification. In: Proceedings of the 2nd Workshop on Speech and Language Processing for Assistive Technologies, pp. 73–83. Association for Computational Linguistics, Stroudsburg (2011)
- Delorme, A., Makeig, S.: EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* **134**(1), 9–21 (2004)
- Delorme, A., Kothe, C., Vankov, A., Bigdely-Shamlo, N., Oostenveld, R., Zander, T.O., Makeig, S.: MATLAB-based tools for BCI research. In: Brain–Computer Interfaces, pp. 241–259. Springer, London (2010)
- DuBay, W.H.: The principles of readability. Online Submission (2004)
- Durkee, K., Hiriyanna, A., Pappada, S., Feeney, J., Galster, S.: Multi-model approach to human functional state estimation. In: International Conference on Augmented Cognition, pp. 188–197. Springer, Cham (2016)
- Feng, L.: Automatic readability assessment. City University of New York (CUNY). Retrieved from <http://lijun.symptotic.com/files/thesis.pdf?attredirects=0> (2010)
- Flesch, R.: A new readability yardstick. *J. Appl. Psychol.* **32**(3), 221 (1948)
- François, T., Fairon, C.: An “AI readability” formula for French as a foreign language. In: Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (2012)

- Friedman, D.B., Hoffman-Goetz, L.: A systematic review of readability and comprehension instruments used for print and web-based cancer information. *Health Educ. Behav.* **33**(3), 352–373 (2006)
- Gannon, E., He, J., Gao, X., Chaparro, B.: RSVP reading on a smart watch. In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 60, no. 1, pp. 1130–1134. SAGE Publications, Los Angeles (2016)
- Gerjets, P., Walter, C., Rosenstiel, W., Bogdan, M., Zander, T.O.: Cognitive state monitoring and the design of adaptive instruction in digital environments: lessons learned from cognitive workload assessment using a passive brain–computer interface approach. *Front. Neurosci.* **8**, 385 (2014). <https://doi.org/10.3389/fnins.2014.00385>
- Gevins, A., Smith, M.E.: Neurophysiological measures of working memory and individual differences in cognitive ability and cognitive style. *Cereb. Cortex* **10**(9), 829–839 (2000). <https://doi.org/10.1093/cercor/10.9.829>
- Gevins, A., Smith, M.E., McEvoy, C.L., Yu, D.: High-resolution EEG mapping of cortical activation related to working memory: effects of task difficulty, type of processing, and practice. *Cereb. Cortex* **7**(4), 374–385 (1997). <https://doi.org/10.1093/cercor/7.4.374>
- Graesser, A.C., McNamara, D.S., Cai, Z., Conley, M., Li, H., Pennebaker, J.: Coh-metrix measures text characteristics at multiple levels of language and discourse. *Elem. School J.* **115**(2), 210–229 (2014)
- Graesser, A., Louwrese, M., McNamara, D., Olney, A., Cai, Z., Mitchell, H.: Inference generation and cohesion in the construction of situation models: some connections with computational linguistics. In: Schmalhofer, F., Perfetti, C.A. (eds.) *Higher Level Language Processes in the Brain: Inference and Comprehension Processes*, pp. 289–310. Erlbaum, Mahwah (2007)
- Grimes, D., Tan, D.S., Hudson, S.E., Shenoy, P., Rao, R.P.: Feasibility and pragmatics of classifying working memory load with an electroencephalograph. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 835–844. ACM (2008)
- Hancke, J., Vajjala, S., Meurers, D.: Readability classification for German using lexical, syntactic, and morphological features. In: *Proceedings of COLING 2012*, December, pp. 1063–1080 (2012)
- Hart, S.G. (2006). NASA-task load index (NASA-TLX); 20 years later. In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 50, no. 9, pp. 904–908. Sage Publications
- Hart, S.G., Staveland, L.E.: Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. *Adv. Psychol.* **52**, 139–183 (1988)
- Haufe, S., Meinecke, F., Görgen, K., Dähne, S., Haynes, J.D., Blankertz, B., Bießmann, F.: On the interpretation of weight vectors of linear models in multivariate neuroimaging. *Neuroimage* **87**, 96–110 (2014)
- Hedin, B., Lindgren, E.: A comparison of presentation methods for reading on mobile phones. *IEEE Distrib. Syst. Online* **8**(6), 2–2 (2007)
- Heilman, M., Collins-Thompson, K., Callan, J., Eskenazi, M.: Combining lexical and grammatical features to improve readability measures for first and second language texts. In: *Human Language Technologies 2007: The Conference of the North American Chapter of the Association for Computational Linguistics (HLT-NAACL-07)*, Rochester, New York, pp. 460–467 (2007)
- Herff, C., Heger, D., Fortmann, O., Hennrich, J., Putze, F., Schultz, T.: Mental workload during n-back task—quantified in the prefrontal cortex using fNIRS. *Front. Hum. Neurosci.* **7**, 935 (2014)
- Hester, M., Werner, S., Greenwald, C., Gunning, J.: Exploring the effects of text length and difficulty on RSVP reading. In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 60, no. 1, pp. 1294–1298. SAGE Publications, Los Angeles (2016)
- Hettinger, L.J., Branco, P., Encarnacao, L.M., Bonato, P.: Neuroadaptive technologies: applying ergonomics to the design of advanced interfaces. *Theor. Issues Ergon. Sci.* **4**(1–2), 220–237 (2003)
- Hudlicka, E., Mcneese, M.D.: Assessment of user affective and belief states for interface adaptation: application to an Air Force pilot task. *User Model. User Adapt. Interact.* **12**(1), 1–47 (2002)
- Iversen, J.R., Makeig, S.: MEG/EEG data analysis using EEGLAB. In: *Magnetoencephalography*, pp. 199–212. Springer, Berlin (2014)
- Jayaram, V., Alamgir, M., Altun, Y., Scholkopf, B., Grosse-Wentrup, M.: Transfer learning in brain-computer interfaces. *IEEE Comput. Intell. Mag.* **11**(1), 20–31 (2016)
- Just, M. A., Carpenter, P. A.: A capacity theory of comprehension: individual differences in working memory. *Psychol. Rev.* **99**(1), 12 (1992)
- Kendeou, P., McMaster, K.L., Christ, T.J.: Reading comprehension: core components and processes. *Policy Insights Behav. Brain Sci.* **3**(1), 62–69 (2016)

- Kincaid, J.P., Fishburne Jr, R.P., Rogers, R.L., Chissom, B.S.: Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel (1975)
- Kintsch, W.: *Comprehension: A Paradigm for Cognition*. Cambridge University Press, New York (1998)
- Kleiner M., Brainard D., Pelli D.: What's new in Psychtoolbox-3? Perception 36 ECVP Abstract Supplement (2007)
- Kothe, C.A., Makeig, S.: BCILAB: a platform for brain-computer interface development. *J. Neural Eng.* **10**(5), 056014 (2013)
- Kramer, A.F.: Physiological metrics of mental workload: a review of recent progress. In: *Multiple-task performance*, pp. 279–328 (1991)
- Krauledat, J.M.: Analysis of nonstationarities in EEG signals for improving brain-computer interface performance (2008)
- Krause, C.M., Sillanmäki, L., Koivisto, M., Saarela, C., Häggqvist, A., Laine, M., Hämäläinen, H.: The effects of memory load on event-related EEG desynchronization and synchronization. *Clin. Neurophysiol.* **111**(11), 2071–2078 (2000). [https://doi.org/10.1016/S1388-2457\(00\)00429-6](https://doi.org/10.1016/S1388-2457(00)00429-6)
- Krol, L.R., Zander, T.O.: Passive BCI-based neuroadaptive systems. In: *Proceedings of the 7th Graz Brain-Computer Interface Conference*, vol. 2017, pp. 248–253 (2017)
- Krol, L.R., Freytag, S.C., Fleck, M., Gramann, K., Zander, T.O.: A task-independent workload classifier for neuroadaptive technology: preliminary data. In: *Proceedings of IEEE International Conference on Systems, Man, and Cybernetics, IEEE SMC*, pp. 003171–003174
- Kuhn, M.R., Schwaneflugel, P.J., Meisinger, E.B.: Aligning theory and assessment of reading fluency: automaticity, prosody, and definitions of fluency. *Read. Res. Q.* **45**(2), 230–251 (2010)
- McNamara, D.S., Kintsch, E., Songer, N.B., Kintsch, W.: Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from text. *Cognit. Instr.* **14**(1), 1–43 (1996)
- Müller-Putz, G., Scherer, R., Brunner, C., Leeb, R., Pfurtscheller, G.: Better than random: a closer look on BCI results. *Int. J. Bioelectromagn.* **10**, 52–55 (2008)
- Moravcsik, J.E., Kintsch, W.: Writing quality, reading skills, and domain knowledge as factors in text comprehension. *Can. J. Exp. Psychol.* **47**(2), 360 (1993)
- Ojala, M., Garriga, G.C.: Permutation tests for studying classifier performance. *J. Mach. Learn. Res.* **11**, 1833–1863 (2010)
- Öquist, G., Goldstein, M.: Towards an improved readability on mobile devices: evaluating adaptive rapid serial visual presentation. *Interact. Comput.* **15**(4), 539–558 (2003)
- Perfetti, C.A.: *Reading Ability*. Oxford University Press, New York (1985)
- Pesonen, M., Hämäläinen, H., Krause, C.M.: Brain oscillatory 4–30 Hz responses during a visual n-back memory task with varying memory load. *Brain Res.* **1138**, 171–177 (2007). <https://doi.org/10.1016/j.brainres.2006.12.076>
- Petersen, S.E., Ostendorf, M.: A machine learning approach to reading level assessment. *Comput. Speech Lang.* **23**, 86–106 (2009)
- Roy, R.N., Charbonnier, S., Campagne, A., Bonnet, S.: Efficient mental workload estimation using task-independent EEG features. *J. Neural Eng.* **13**(2), 026019 (2016)
- Salomon, G.: Television is "easy" and print is "tough": the differential investment of mental effort in learning as a function of perceptions and attributions. *J. Educ. Psychol.* **76**(4), 647 (1984)
- Scharinger, C., Kammerer, Y., Gerjets, P.: Pupil dilation and EEG alpha frequency band power reveal load on executive functions for link-selection processes during text reading. *PLoS ONE* **10**(6), e0130608 (2015). <https://doi.org/10.1371/journal.pone.0130608>
- Scharinger, C., Soutschek, A., Schubert, T., Gerjets, P.: Comparison of the working memory load in n-back and working memory span tasks by means of EEG frequency band power and P300 amplitude. *Front. Hum. Neurosci.* **11**, 6 (2017). <https://doi.org/10.3389/fnhum.2017.00006>
- Schultheis, H., Jameson, A.: Assessing cognitive load in adaptive hypermedia systems: physiological and behavioral methods. In: *International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems*, pp. 225–234. Springer, Berlin (2004)
- Schwarm, S.E., Ostendorf, M.: Reading level assessment using support vector machines and statistical language models. In: *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics - ACL'05*, pp. 523–530. Association for Computational Linguistics, Morristown (2005). <http://doi.org/10.3115/1219840.1219905>

- Si, L., Callan, J.: A statistical model for scientific readability. In: Proceedings of the 10th International Conference on Information and Knowledge Management (CIKM), pp. 574–576. ACM (2001)
- Sweller, J., van Merriënboer, J.J.G., Paas, F.G.W.C.: Cognitive architecture and instructional design. *Educ. Psychol. Rev.* **10**(3), 251–296 (1998). <https://doi.org/10.1023/A:102219372820>
- Tibus, M., Heier, A., Schwan, S.: Do films make you learn? Inference processes in expository film comprehension. *J. Educ. Psychol.* **105**, 329–340 (2013)
- Vajjala, S.: Analyzing text complexity and text simplification: connecting linguistics, processing and educational applications. Doctoral dissertation, Ph. D. thesis, University of Tübingen (2015)
- Vajjala, S., Meurers, D.: Readability assessment for text simplification: from analyzing documents to identifying sentential simplifications. *Int. J. Appl. Linguist.* **165**, 194–222 (2014)
- Vapnik, V.N.: An overview of statistical learning theory. *IEEE Trans. Neural Netw.* **10**(5), 988–999 (1999)
- Vygotsky, L.S.: *Mind in Society: The Development of Higher Psychological Processes*. Harvard University Press, Harvard (1980)
- Wallot, S., O'Brien, B.A., Haussmann, A., Kloos, H., Lyby, M.S.: The role of reading time complexity and reading speed in text comprehension. *J. Exp. Psychol. Learn. Mem. Cogn.* **40**(6), 1745 (2014)
- Walter, C., Rosenstiel, W., Bogdan, M., Gerjets, P., Spüler, M.: Online EEG-based workload adaptation of an arithmetic learning environment. *Front. Hum. Neurosci.* **11**, 286 (2017). <https://doi.org/10.3389/fnhum.2017.00286>
- Wickens, C.D.: Multiple resources and mental workload. *Hum. Factors* **50**(3), 449–455 (2008)
- Zander, T.O.: Utilizing brain-computer interfaces for human-machine systems. Doctoral dissertation, Universitätsbibliothek der Technischen Universität Berlin (2011)
- Zander, T.O., Kothe, C.: Towards passive brain-computer interfaces: applying brain-computer interface technology to human-machine systems in general. *J. Neural Eng.* **8**(2), 025005 (2011)
- Zander, T.O., Kothe, C., Jatzev, S., Gaertner, M.: Enhancing human-computer interaction with input from active and passive brain-computer interfaces. In: *Brain-Computer Interfaces*, pp. 181–199. Springer, London (2010)
- Zander, T.O., Krol, L.R., Birbaumer, N.P., Gramann, K.: Neuroadaptive technology enables implicit cursor control based on medial prefrontal cortex activity. *Proc. Natl. Acad. Sci.* **113**(52), 14898–14903 (2016)
- Zhang, X., Krol, L.R., Zander, T.O.: Towards task-independent workload classification: shifting from binary to continuous classification. In: *Proceedings of IEEE International Conference on Systems, Man, and Cybernetics. IEEE SMC 2018* (2018)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Lena M. Andreessen has finished her M.Sc. in Human Factors at Technical University Berlin in 2018, where she has continued her interdisciplinary research interests after completing her B.Sc. in Human-Computer Interaction at Universität Hamburg in 2013. Following her interests in Human-Machine Systems, neurophysiological measurement methods and passive Brain-Computer Interfaces, she now is a doctoral candidate at the department for Neuroadaptive Human-Computer Interaction at Brandenburg Technical University, Germany.

Peter Gerjets finished his diploma in psychology at the University of Göttingen in 1991. From 1991 to 1995 he was a Research Associate at the University of Göttingen where he received his Ph.D. in 1994. Afterwards he has been working as Assistant Professor at the Saarland University in Saarbrücken where he finished his habilitation in 2002 before taking over his current position at the University of Tübingen. Since 2002 he has been working as principal research scientist at the Knowledge Media Research Center and besides as full professor for research on learning and instruction at the University of Tübingen. He was honoured with the Young Scientist Award of the German Cognitive Science Society in 1999 and served in the editorial boards of the *Journal of Educational Psychology*, *Educational Research Review*, *Computers in Human Behavior*, *Metacognition and Learning*, and *Educational Technology, Research, and Development*. His current research focuses on multimodal and embodied interaction with digital media

as well as on learning from multimedia, hypermedia, and the Web. He is a member of DGPs, APS, and EARLI and served as coordinator of the EARLI Special Interest Group 6: Instructional Design.

Detmar Meurers is Professor of Computational Linguistics at the University of Tübingen, Germany and on the steering board of the LEAD Graduate School and Research Network in Empirical Educational Science there. As head of the ICALL-Research.com group, his work focuses on Intelligent Computer-Assisted Language Learning, and computational linguistic methods in Second Language Acquisition research and language teaching. He has published on intelligent language tutoring systems, automatic short answer assessment, the automatic analysis of learner corpora, and input enrichment and enhancement applications for language learners.

Thorsten O. Zander is full professor for Neuroadaptive Human-Computer Interaction at Brandenburg University of Technology, Germany, and the founder and leader of Zander Laboratories, Amsterdam, The Netherlands. Since 2019 he is visiting professor at Higher School for Economics, Moscow, Russia. Thorsten is known for defining the concept of Passive Brain-Computer Interfaces and its applications in Neuroadaptive Technologies.

Affiliations

Lena M. Andreessen¹  · **Peter Gerjets**^{2,3} · **Detmar Meurers**³ · **Thorsten O. Zander**^{1,4}

✉ Lena M. Andreessen
lena.andreessen@gmail.com

¹ Brandenburg University of Technology Cottbus-Senftenberg, Cottbus, Germany

² Leibniz-Institut für Wissensmedien, Tuebingen, Germany

³ University of Tuebingen, Tuebingen, Germany

⁴ Center for Bioelectric Interfaces, National Research University Higher School of Economics, Moscow, Russia