International Journal of Psychophysiology 158 (2020) 201–214

Contents lists available at ScienceDirect

International Journal of Psychophysiology

journal homepage: www.elsevier.com/locate/ijpsycho

# Using eye-tracking and EEG to study the mental processing demands during learning of text-picture combinations

Christian Scharinger[a,*], Anne Schüler[a], Peter Gerjets[a,b]

[a] Leibniz-Institut für Wissensmedien Tübingen, Germany
[b] Department of Cognitive Psychology and Media Psychology, University of Tübingen, Germany

ARTICLE INFO

*Keywords:*
EEG
Alpha/theta frequency band power ERD/ERS%
Fixation-related EEG data analysis
Eye-tracking
Pupil dilation
Multimedia
Text-picture integration
Mental processing demands

ABSTRACT

Using and combining eye-tracking and EEG frequency band power as process measures, in the current study we were interested in the mental processing demands during learning of text-picture combinations that either enabled or prohibited text-picture integration (TPI). In the mismatch condition, the textual and pictorial information being dissimilar, TPI was not possible, whereas in the match and the partial-match condition, the textual and pictorial information being identical respective complementary, TPI was possible. We expected mental processing demands to be higher in the mismatch condition, when pictorial and textual information had to be processed and memorized as separate representations, compared to the match and partial-match conditions when TPI was possible. As expected, on virtually all process measures we observed increased mental processing demands when two mental representations had to be processed and memorized compared to the two conditions where TPI was possible. The EEG alpha and theta frequency band power data corroborated and extended the eye-tracking measures of mental processing demands. In addition, we performed a fixation-related EEG frequency band power analysis that also corroborated the results of the classic stimulus-locked EEG frequency band power analysis, exemplifying the use of this former methodology in the context of complex multimedia task materials.

## 1. Introduction

Identifying adequate process measures to study and better understand cognitive effects of instructional design decisions has become a central focus in instructional psychology during the last two decades (Harteis et al., 2018). Especially, eye-tracking has become increasingly popular for analyzing instructional materials (e.g., van Gog and Scheiter, 2010). Furthermore, other physiological measures like the electroencephalography (EEG) and, more specifically, EEG frequency band power, have been proposed as potentially promising process measures (Antonenko et al., 2010).

In the current study, we used and combined eye-tracking and EEG to study text-picture integration (TPI). TPI has been proposed to be the most pivotal process when learning with multimedia materials, as the resulting integrated model allows for deeper understanding (Mayer, 2009; Schnotz, 2014). To study TPI, we assessed the mental processing demands associated with three different types of text-picture combinations: First, text-picture combinations with both representations conveying the same information (i.e., identical information; termed *match* condition henceforth). Second, text-picture combinations with both representations conveying complementary information, that is, both representations depicted respective described the same content but one representation contained one piece of information that was more specific (termed *partial-match* condition henceforth). Third, text-picture combinations with both representations being unrelated (i.e., containing different information; termed *mismatch* condition henceforth). Importantly, whereas in the match and partial-match condition it was possible to integrate text and picture with each other (cf. Arndt et al., 2015; Schüler et al., 2015, 2019), this was not possible in the mismatch condition. Instead, in the mismatch condition learners had to process and memorize both representations individually. Hence, finding differences in the mental demands associated with the conditions where TPI was possible and the condition where TPI was not possible, would be a strong indicator that text and pictures allowing for TPI are indeed processed differently than text and pictures which do not allow for TPI.

As we were specifically interested in the differences in mental processing demands induced *during* learning of text-picture combinations when TPI was not possible as compared to TPI being possible, in the current study we focused on eye-tracking and EEG measures that have been reported in the literature to specifically reflect mental processing

* Corresponding author at: Leibniz-Institut für Wissensmedien Tübingen, Schleichstr. 6, 72076 Tübingen, Germany.
*E-mail address:* c.scharinger@iwm-tuebingen.de (C. Scharinger).

https://doi.org/10.1016/j.ijpsycho.2020.09.014
Received 5 March 2020; Received in revised form 10 September 2020; Accepted 27 September 2020
Available online 17 October 2020

demands. In detail these measures were for the eye-tracking data the pupil dilation (e.g., Kahneman and Beatty, 1966; Mathôt, 2018). For the EEG data these measures were the EEG theta (4–6 Hz) frequency band power at frontal electrodes and the EEG upper alpha (10–13 Hz) frequency band power at parietal electrodes (e.g., Brouwer et al., 2014; Fairclough and Ewing, 2016; Gevins and Smith, 2000; Palomäki et al., 2012; Scharinger et al., 2015b).

Furthermore, analyzing eye-tracking data and EEG data simultaneously in one study was thought to provide further insights into the sensitivity of these measures to assess mental processing demands in the context of complex task materials (and the feasibility of using these measures therein). Finally, the combined recording of eye-tracking and EEG data allowed us to additionally analyzing the EEG data fixation-related, that is, when certain areas of interest (AOIs) of the task materials are fixated at (e.g., the text or the picture), which might be in general a promising methodology for studying the EEG in complex, multimedia task materials (Scharinger, 2018).

### 1.1. Text-picture integration (TPI)

When learning textual information, adding pictures to text (i.e., creating multimedia presentations) has been shown to be beneficial for learning (Anglin et al., 2004; Butcher, 2014). This beneficial effect of adding picture to text has been explained by the assumption that text and pictures are integrated mentally with each other into one coherent mental model. This integrated mental model, in turn, allows for deeper understanding, resulting in better learning outcomes. These theoretical assumptions are grounded in the 'cognitive theory of multimedia learning' (CTML; Mayer, 2009), and in the 'integrated model of text and picture comprehension' (Schnotz, 2014). For example, CTML assumes that learners construct a mental representation of the text (i.e., a *verbal* mental model, as it is traditionally termed in instructional psychology, irrespective whether the text is written or spoken) and a mental representation of the picture (i.e., a pictorial mental model) in working memory (Baddeley and Hitch, 1974; Baddeley, 2012). Note that in the following, we will use the term *textual* instead of *verbal* to clarify that in the current study we focused on visual task materials only (i.e., written sentences and pictures). In a next step, learners integrate the textual and the pictorial mental representations with each other, resulting in one coherent mental model. According to Mayer (2009) this process of TPI requires the use of working memory resources. It involves building one-to-one correspondences between the textual and the pictorial mental representation. For example, a learner may try to connect the textual mental representation to the pictorial mental representation by looking for how a sentence in the text corresponds to a part of the picture. Mayer assumes that TPI is the most crucial step in multimedia learning, because it will not only support remembering the information presented, but also foster deeper understanding of the contents. Of course, a prerequisite for TPI to occur is that the textual and pictorial information are content-related (i.e., are identical or contain the same plus complementary information), as only then the textual and pictorial information can be referred to each other and potentially be integrated in memory, whereas for content-unrelated textual and pictorial information, integration is not possible and two separate mental representations (one pictorial, one textual) have to be stored in memory.

Empirical research addressing TPI in multimedia learning materials often focuses on outcome measures, for example reporting enhanced transfer performance due to TPI (Mayer, 2009). Schüler and colleagues (Arndt et al., 2015; Schüler et al., 2015, 2019) reported the performance in a sentence-verification task after multimedia learning as a more direct hint for TPI. In the multimedia learning task subjects had to memorize text-picture combinations of single written sentences combined with simple line-drawings that described respective depicted easily identifiable actions or situations. Text and picture were either presented in a general or in a specific version, with the information provided in the text and in the picture being either identical (i.e., general or specific information provided in both representations) or complementary (i.e., specific information provided in only one of the representations, for example the picture, and general information provided in the other representation, for example the text). The authors observed that in complementary conditions subjects more often falsely recognized the specific version of a sentence when they had previously memorized the general version of that sentence combined with a specific version of the corresponding picture. The authors interpreted these recognition biases as a strong hint for TPI. Yet, a drawback of these multimedia learning studies lies in the fact that they addressed TPI only after the learning process has happened. Therefore, process measures like the EEG are promising as they allow studying effects of TPI directly *during* learning.

There are several EEG studies that aimed at more directly assessing the cognitive processes related to TPI (e.g., Coppens et al., 2012; Knoeferle et al., 2011; Li et al., 2020). Yet, these EEG studies predominantly focus on event-related potential (ERPs) and not on EEG frequency band power analyses for studying the cognitive effects of processing textual and pictorial information. Furthermore, these studies seldom use complex multimedia task materials or true learning paradigms. More importantly for the current study, to the best of our knowledge, to date none of these EEG studies directly addressed and examined the effects of TPI on mental processing demands using EEG frequency band power data. This is somewhat surprising, as TPI in theory (cf. above) has been closely related to working memory and is hypothesized to alter the mental processing demands. By analyzing EEG and eye-tracking measures of mental processing demands the current study aims at closing this gap in research.

### 1.2. Physiological process measures of mental processing demands

For the eye-tracking data, especially the pupil diameter has been reported as a sensitive measure of task demands, at least for younger adults and if the luminance of the multimedia materials is sufficiently controlled for (Cabestrero et al., 2009; Kahneman and Beatty, 1966; Laeng et al., 2012; Mathôt, 2018; van der Wel and van Steenbergen, 2018; Van Gerven et al., 2004). Typically, the pupil dilates for increased mental processing demands.

The EEG frequency band power, and more specifically the EEG theta (4–6 Hz) frequency band power at frontal electrodes and the EEG upper alpha (10–13 Hz) frequency band power at parietal electrodes have been shown to be sensitive for measuring mental processing demands (e.g., due to increased working memory load) in a variety of different tasks, so far predominantly in highly controlled task settings of basic research (e.g., Brouwer et al., 2014; Fairclough and Ewing, 2016; Gevins and Smith, 2000; Wolfgang Klimesch, 1999; Palomäki et al., 2012; Scharinger et al., 2015b). Increased mental processing demands typically result in increased EEG theta power at frontal electrodes and decreased (upper) alpha power at parietal electrodes. A (relative) increase in frequency band power is typically termed event-related synchronization (ERS), a (relative) decrease in frequency band power is termed event-related desynchronization (ERD). One prominent way to express changes in EEG frequency band power (with respect to a baseline) is the use of the ERD/ERS%-measure (Pfurtscheller and Lopes da Silva, 1999). ERD/ERS%-values are calculated using the following formula: ERD/ERS% = 100 * (frequency band power condition − frequency band power baseline) / frequency band power baseline (Cohen, 2014; Antonenko et al., 2010; Pfurtscheller and Lopes da Silva, 1999). In principal, three different baseline-selections are possible for calculating ERD/ERS%-values: a) using a pre-stimulus baseline, b) using a separate condition as baseline (e.g., a rest baseline), or c) using a whole trial condition-average baseline (Cohen, 2014). Note that the ERD/ERS%-values are calculated separately for each electrode and each frequency band. Depending on the concrete selection of the baseline for calculating ERD/ERS%-values an ERS (or ERD) might not always be associated with a positive (or negative) sign as originally defined by Pfurtscheller and Lopes da Silva (1999).

While both, the EEG upper alpha and the theta frequency band power seem to reflect mental processing demands, the upper alpha frequency band power might especially be related to semantic processing and long-term memory recall (Klimesch et al., 1997a; Klimesch et al., 1997b; Klimesch, 1999, 2012), whereas the theta frequency band power might especially be related to processes of cognitive control and working memory (Cavanagh and Frank, 2014; Sauseng et al., 2010; Sauseng et al., 2005). Importantly, in the context of purely working memory tasks, for purely stimuli-retention, the EEG alpha power has been observed to increase (i.e., showing an alpha ERS instead of an alpha ERD) for increased working memory load (e.g., Jensen et al., 2002). It has been hypothesized that the alpha ERS might reflect internal inhibitory processes (Klimesch et al., 2007; Klimesch, 2012; Palva and Palva, 2007). Taken together, the cognitive processes reflected in the EEG theta and (upper) alpha frequency bands may be diverse and rather elusive (for reviews see e.g., Klimesch, 1999; Krause, 2003), yet some functional differentiations between theta frequency band power effects and upper alpha frequency band power effects seem to be plausible, and, more importantly for the current study, both measures seem to sensitively react to mental processing demands.

Noteworthy, although most research on EEG frequency band power has been done in the context of highly controlled tasks of basic research, during the last years there have also been some promising (yet rare) studies that used the EEG alpha and theta ERD/ERS%-measure in the context of complex instructional materials, for example to study the effects of leads (i.e., short previews) on hypertext reading (Antonenko and Niederhauser, 2010), hyperlink-selection processes (Scharinger et al., 2015a), or the spatial contiguity effect in multimedia learning (Makransky et al., 2019). To the best of our knowledge, the EEG upper alpha and theta frequency band power have not been used before for studying TPI and, moreover, especially not in combination with eye-tracking as it has been done in the current study.

A longstanding assumption for eye-tracking data is that eye fixations signal the processing of the fixated visual content (Hyönä, 2010; Just and Carpenter, 1980; Rayner, 2009). Thus, analyzing the EEG frequency band power fixation-related (i.e., time-locked to specific fixations, for example the first fixation on the text after the picture had been looked at) may provide further insights in the timing of mental processing demands. In addition, we conducted such an analysis in the current study exploratorily in addition to the classical stimulus-onset locked EEG analysis to verify the validity of the fixation-related EEG frequency band power analysis (i.e., expecting that both analyses would result in comparable outcomes).

### *1.3. The present study*

In the current study we were interested in investigating TPI during learning by using eye-tracking and the EEG as process measures and, furthermore, by combining both measure with another. Importantly, we focused on measures like the EEG alpha and theta frequency band power and the pupil dilation, that are specifically indicative for mental processing demands.

We compared the mental processing demands of three conditions with each other: First, the match condition, where both representations (i.e., text and picture) conveyed the same information (i.e., allowing for TPI). Second, the partial-match condition, where both representations conveyed mainly the same information, with the exception that one representation contained one piece of information that was more specific (i.e., still allowing for TPI). And third, the mismatch condition, where both representations were unrelated (i.e., making TPI impossible). We hypothesized that if integration occurs during learning, mental processing demands should differ between conditions where integration is possible and the condition where integration is not possible. More specifically, we expected higher processing demands in the mismatch condition than in the other two conditions, because in the mismatch condition two separate mental representations (i.e., one textual and one pictorial) had to be constructed and memorized, whereas in the match and partial-match condition two separate mental representations had to be constructed, which then could be integrated and memorized as one integrated model. Regarding the differences of mental processing demands in the two conditions where integration was possible, two outcomes were imaginable in our view: First, no differences regarding mental processing demands, as in both conditions TPI was possible. Second, higher mental processing demands in the partial-match condition as one representation contained complementary information which had to be integrated into the resulting mental model.

As a manipulation check we added a sentence-verification test. It allowed us to verify on a performance level that learners in the integration conditions indeed integrated text and pictures with each other: In line with the studies conducted by Schüler and colleagues (e.g., Arndt et al., 2015; Schüler et al., 2015), we expected less accuracy in the sentence-verification task for the partial-match condition compared to the match condition, because when participants had previously memorized the general version of a sentence combined with a specific version of the corresponding picture this should lead to falsely recognizing the specific version of a sentence in the sentence-verification test.

Exploratorily, in the current study we wanted to compare the outcomes of a classical stimulus-locked EEG frequency band power analysis with the outcomes of a fixation-related EEG frequency power analysis, and hence the validity of the latter one for studying multimedia materials. Fixation-related EEG data analysis has been proposed as a promising methodology to study the EEG in free viewing situations of text or multimedia materials, yet most studies so far focused on the time-domain (e.g., Dimigen et al., 2011), that is on fixation-related potentials and not on the frequency domain, that is, fixation-related frequency band power analysis (but see Scharinger et al., 2015a). Thus, the current study intends to additionally examine the potentials of fixation-related EEG frequency band power analysis for studying mental processing demands in multimedia materials.

## 2. Method

### *2.1. Participants*

Twenty-five healthy subjects (university students, mean age = 23.04, *SD* = 4.15, 19 females) participated in the study for a payment of 8 € per hour. They were all native speakers of German, right-handed as indicated by the Edinburgh Handedness Inventory (Oldfield, 1971), and had normal or corrected-to-normal visual acuity. The study has been approved by the local ethics committee and was conducted in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki and its later amendments or comparable ethical standards. All participants gave their written informed consent prior to their inclusion in the study.

### *2.2. Stimuli*

As stimuli served text-picture combinations consisting of simple black-and-white line drawings with a one-line sentence below each drawing (see Fig. 1). All line-drawings were presented within a thin black frame of equal size (860 × 650 px) that was horizontally centered on the screen and vertically aligned with a distance of 56 px to the upper border of the screen. The line-drawings span a visual angle of 19.50° horizontally and 15.16° vertically. The technical brightness of the line-drawings was on average 198.93 (*SD* = 6.36; RGB scale 0 to 255). The sentence (Arial, 30 pt) was positioned centrally below the line-drawing (with a distance of 80 px) and consisted of four to nine words. The sentences span a visual angle of about 8.46° to 21.94° horizontally (depending on sentence length) and 1.49° vertically.

The stimuli were selected out of a pool of text-picture combinations that had been used in previous studies (cf. Arndt et al., 2015; Schüler et al., 2015; see Fig. 1). The pool consisted of 138 pictures (simple line-

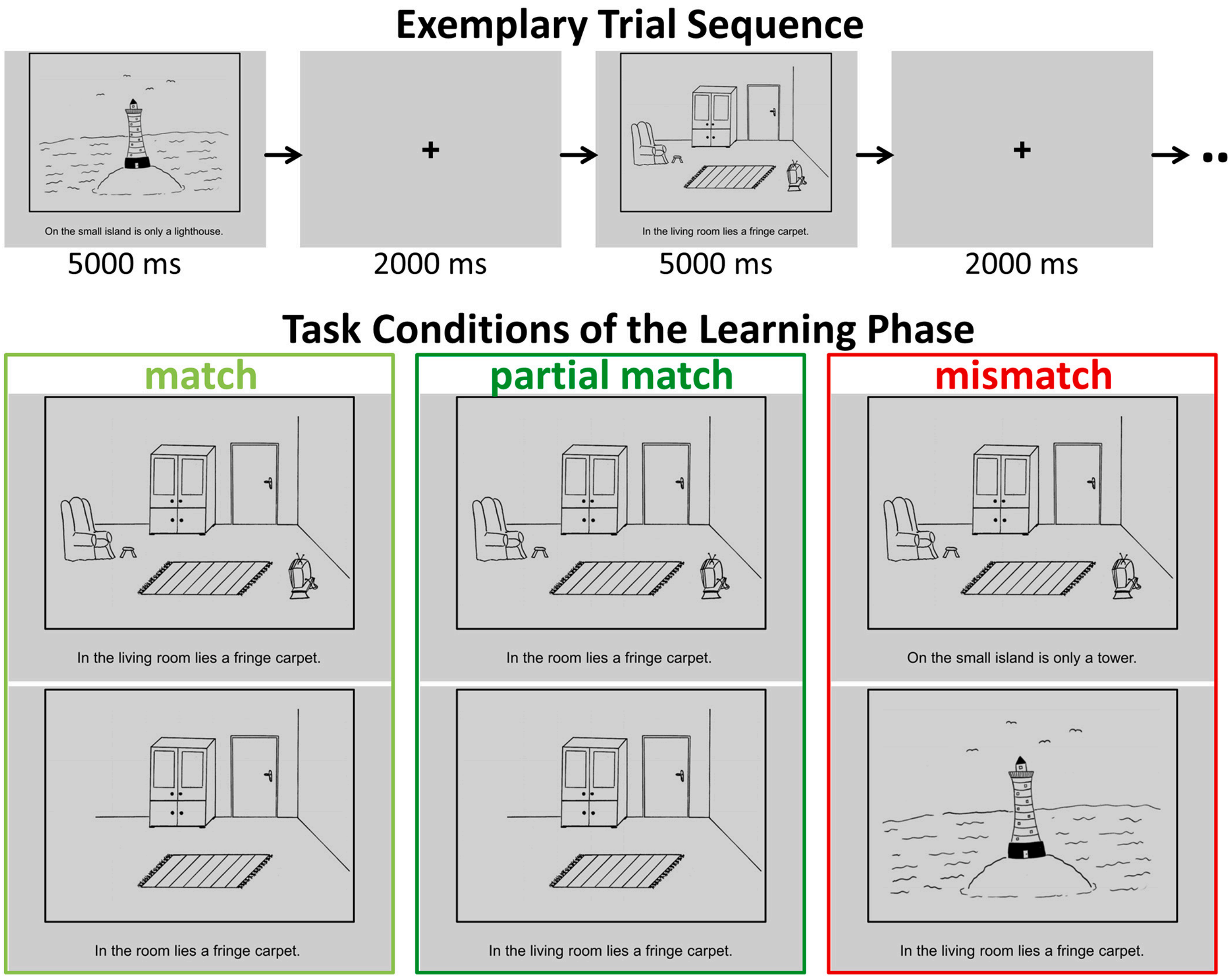


**Fig. 1.** Exemplary trial sequence and task conditions of the learning phase. In the match condition the information content of text and picture was identical. In the partial-match condition text and picture were complementary. In the mismatch condition the text and the picture referred to unrelated information content. Note. The size ratio of text and picture presented here has been changed for reasons of space limitation and does not match the size ratio of the original task materials. The language of the original task materials was German.

drawings) and 138 sentences. Half of the pictures (i.e., 69) depicted situations or objects that were of general information content (e.g., a fringed carpet lying in a neutral room), half of the pictures depicted situations or objects that were of specific information content. Importantly, the specific version of a picture was generated by adding specific features to the general version of the picture showing the same topic (e.g., the specific version of the example above would be a fringed carpet lying in a specific room, namely a living room as indicated by additional elements like a TV set and an armchair). The sentences in the pool were descriptions of the depicted content in the pictures. Thus, comparably to the pictures, half of the sentences (i.e., 69) of the pool described a general situation (e.g., "In the *room* lies a fringed carpet."[1]), half of the sentences described a specific situation (e.g., "In the *living room* lies a fringed carpet."). The specific sentence-version was created out of the general sentence-version of the same topic by specifying the object or situation described more closely.

Out of this pool of sentences and pictures the stimuli for three different conditions with respect to TPI were created: In the match condition (TPI possible), a line-drawing was combined with a corresponding sentence that was of the same specificity (i.e., general/general or specific/specific). In the partial-match condition (TPI possible) the line-drawing and the sentence were complementary (i.e., either a specific text combined with a general picture or vice versa). In the mismatch condition (TPI not possible) the depicted situation in the picture and the described situation in the text were completely different.

### 2.3. Procedure

The study consisted of a learning phase and a recall phase. In the learning phase, the text-picture combinations were presented sequentially in five blocks. Participants were instructed to memorize each text and each picture for later recall. They were informed that the learning phase consisted of five blocks with a recall phase after the fifth block. However, participants were not informed beforehand how exactly the recall test would look like. Each block consisted of 48 trials, one trial consisting of a text-picture combination that was presented for 5000 ms followed by a central fixation-cross presented for 2000 ms (cf. Fig. 1). Consequently, in sum one trial lasted 7000 ms, one block lasted about 5 ½ min. One third ($n = 16$) of the trials of a block were matches, one third ($n = 16$) partial-matches and one third ($n = 16$) mismatches. Over all

[1] Note. Literal translation of the stimulus materials that were originally presented in German.

five blocks, there were 80 trials per condition. The sequence of trials (i.e., conditions) in a block was completely random.

For each participant the stimuli of the three conditions were individually created by randomly choosing textual and pictorial materials out of the pool of textual and pictorial materials described above (Section 2.2) with the following constraints: within a block, one depicted or described topic could occur only once. However, the same picture or sentence could be repeated in another block, yet not in the identical combination (i.e., over all blocks one specific combination of text and picture could occur only once). Overall, the number of general and specific text-picture combinations was equally distributed across the three conditions. Therefore, potentially confounding visual factors like the sentence length or the pictorial density could be ruled out to differently affecting the three conditions. Note that the repetition of a picture or sentence (yet in another combination) during the course of the study was an inevitable compromise enabling us to use the same task materials that have been successfully used before in studies on TPI (Arndt et al., 2015; Schüler et al., 2015) while ensuring to have enough stimuli per task condition for reliable EEG data analyses. Note additionally, that as a consequence of this procedure in the first block all presented content topics were new to the participants whereas in the following four blocks only few additional new content topics were introduced and most content topics reoccurred, yet in different task conditions (i.e., different text-picture combinations).

Between the blocks participants had short breaks. After the five blocks of the learning phase, the recall phase started. A sentence-verification test was administered therein, which was explicitly connected only to the stimulus materials presented in the fifth block of the previous learning phase to test learning success. The sentence-verification test consisted of 96 sentences. The sentences were presented at a fixed timing of 2000 ms each followed by a centrally positioned fixation-cross for 500 ms. Half of the sentences had been presented in the directly preceding block of the learning phase (i.e., block 5), half of the sentences had been presented in the other four blocks. Participants were instructed to decide as correctly and as quickly as possible binary via key-pressing whether the sentence had been part of the directly preceding block 5 of the learning phase or not. We calculated the accuracy-score for yes-responses only, that is, for the 48 sentences that had been presented before in block 5, as the percentage of correct yes-responses for these sentences in relation to the potentially absolute number of yes-responses. For this score, we statistically analyzed whether the accuracy differed between the three conditions (i.e., match, partial-match, and mismatch). The main purpose of the sentence-verification test was to check whether participants performed the learning phase as instructed. Accuracies above chance-level in the match and the mismatch condition (i.e., > 50% correct) should indicate a successful learning phase. Furthermore, it allowed us to exploratorily analyze whether participants in the partial-match condition integrated text and pictures with each other as described by Arndt et al. (2015) and Schüler et al. (2015, 2019). Because of reasons of time and to avoid potentially confounding testing-effects, the sentence-verification task was only presented once after the last block of the learning phase (note that in the current study our primary focus was on the mental processing demands during the learning phase; for effects of TPI on learning outcomes see Arndt et al., 2015 and Schüler et al., 2015, 2019).

### 2.4. Apparatus

The study was run in a dimly lit, quiet room. Participants sat in a comfortable chair in front of a 22-in. Dell monitor (1680 × 1050 pixels screen resolution). A 250 Hz SMI remote eye-tracking system below the monitor recorded participants' gaze behavior at a sampling rate of 250 Hz (SMI iView X 2.7.13). A chin rest was used to avoid head movements during data recording and to ensure the eyes remained a fixed distance of about 70 cm from the eye tracking device. The eye-tracker was calibrated at the beginning and after each break using the built in calibration routines (SMI Experiment Center, 9-point calibration).

EEG was recorded at 28 electrode sites (Fp1, Fp2, F7, F3, Fz, F4, F8, FC5, FC1, FC2, FC6, T7, C3, Cz, C4, T8, TP9, CP5, CP1, CP2, CP6, P7, P3, Pz, P4, P8, O1, O2) positioned according to the international 10/20 system (Jasper, 1958). The right mastoid served as reference during recording. The ground electrode was positioned at AFz. Three additional electrodes were placed around the eyes for the recording of the vertical and horizontal electrooculogram (EOG). EEG data were recorded (PyCorder 1.0.2) at 500 Hz sampling rate (ActiCHamp, Brainproducts, Inc.) using active electrodes (ActiCap, Brainproducts, Inc.). Impedances were kept below 5 kΩ for the EEG electrodes.

### 2.5. Data preprocessing and analysis

We conducted two separate EEG data analyses, one based on data epochs time-locked to stimulus-onset (i.e., a classical stimulus-locked EEG data analysis), and a second one based on data epochs time-locked to certain eye-fixations (i.e., a fixation-related EEG data analysis). Both data analyses share however some common preprocessing steps detailed below.

The eye-tracking data and the EEG data were synchronized and preprocessed using customized Matlab scripts (Matlab 2018b, MathWorks, Inc.) and the toolbox EEGLAB (v. 14.1.2; Delorme and Makeig, 2004) with the EYE-EEG plugin (v.0.85; Dimigen et al., 2011). Event markers at each stimulus onset in both, the EEG data and the eye-tracking data, served as synchronization events. Overall, the average jitter between the markers in the EEG and the corresponding markers in the eye-tracking data was within ±1 data sample (i.e., ±2 ms). The eye-tracking data was integrated in the EEG data as additional channels (e.g., channels that contain the pupil dilation data of each eye or the raw gaze positions) with the original sampling rate of the eye-tracking data of 250 Hz upsampled to 500 Hz to match the sampling rate of the EEG data. Based on the algorithm implemented in EYE-EEG for saccades detection, fixations were identified and stored as events along with the EEG data. These identified fixations were later used for the fixation-related EEG data analysis (see below). Zero-values in the channels containing the pupil dilation data (i.e., indicating blinks) were interpolated using an algorithm and corresponding Matlab code by Greg Siegle (Siegle et al., 2003).

The continuous EEG data were filtered (low-pass 48 Hz, high-pass 0.25 Hz, linear finite impulse response filters). EOG-artifacts (eye-movements, blinks) were corrected for by using independent component analysis (ICA) decompositions. Independent components (ICs) visually identified as EOG-ICs were rejected (Delorme et al., 2007; Jung et al., 2000). The EEG data then were re-referenced to average reference (the EOG-channels as well as the two electrodes positioned on the mastoids [TP9, TP10] were excluded, resulting in 27 symmetrically distributed EEG channels over the scalp that were in the final data set used for statistical analyses).

For the classical, stimulus-onset based EEG data analysis, the continuous EEG data were divided in stimulus-locked epochs of 7500 ms length (i.e., from −250 ms pre-stimulus onset to 7250 ms post-stimulus onset). Using these epochs, an automatic artifact removal was performed: Epochs that exceeded ±100 μV were excluded from further analyses (Duncan et al., 2009; Pesonen et al., 2007). No further artifact removal or correction was performed on the EEG data. In the final data set we had on average 79.16 ($SD = 1.40$) artifact-free trials in the match condition, 78.64 ($SD = 1.93$) in the partial-match, and 78.64 ($SD = 2.64$) artifact-free trials in the mismatch condition. As revealed by a one-factorial repeated-measures ANOVA the number of trials per condition was statistically not different ($F < 1.97$, $p > .150$).

#### 2.5.1. Stimulus-locked EEG data analyses

In the stimulus-locked analysis, we first calculated time-frequency representations (TFRs) of the EEG data (Pesonen et al., 2007), that is, representations of the EEG power expressed as percentage of event-related desynchronization and synchronization (ERD/ERS%; Pfurtscheller and Lopes da Silva, 1999) as a function of time and frequency in the same matrix. The calculations were performed for each EEG channel separately. Using a Morlet wavelet (width 8, i.e., the tapering Gaussian consisting of 8 cycles constantly for all frequencies; cf. Pesonen et al., 2007) TFRs were calculated for the EEG frequencies 4–30 Hz for each participant and each of the three conditions, averaged over corresponding trials.[2] We then calculated the ERD/ERS%-values. The ERD/ERS% gives the percentage of change in EEG frequency band power in a task condition of interest compared to a certain baseline. As baseline we used a 300 ms time-window during the presentation of the fixation-cross situated 200 ms before the onset of the following stimulus (Cohen, 2014), that is, (except for the first trial) a pre-stimulus time-window with minimal and stable visual input. The raw power values of the baseline were averaged over time, trials, and conditions (i.e., forming a global baseline for all conditions) and then used to calculate the ERD/ERS%-values of the TFRs. The grand-average ERD/ERS%-TFRs of electrode Fz and Pz are given in Fig. 4. For further statistical analyses, we averaged the ERD/ERS%-values of the TFRs separately for two frequency bands (theta, 6–8 Hz, upper alpha, 10–13 Hz) and two time-epochs, the first epoch ranging from 0 to 5000 ms in the time-domain, reflecting the time-window of the stimulus presentation, and the second epoch ranging from 5000 to 6500 ms in the time-domain, reflecting the presentation of the fixation cross after the stimulus. Note that the second epoch excluded the time-window used as baseline for calculating the ERD/ERS%-values.

We decided to basically conduct two separate statistical analyses. First, based on literature (e.g., Gevins and Smith, 2000) we selected two typical electrodes, namely the frontal-midline electrode Fz for the analysis of the theta frequency band ERD/ERS% and the parietal-midline Pz for the analysis of the upper alpha frequency band ERD/ERS%. For each of these two frequency bands we conducted separate two-factorial within-subject repeated-measures ANOVAs with the factors *time* (first, second time-window) and *condition* (match, partial-match, mismatch).

We included the analysis of the second time-window covering the post-stimulus fixation-cross for two reasons. First, the fixation-cross presented centrally on the screen provided a time-period with identical visual input for all three task conditions, and, more importantly, with virtually no eye-movements. Thus, especially for the analysis of the pupil dilation data such a time-period with very reduced, stable visual input seemed to be ideally suited. Second, we expected potential TPI processes to spill over into the time-window of the fixation cross, that is, we expected that during the presentation of the fixation-cross processes of (potential) TPI and especially the memorization of the previously processed visual input would continue to occur. Furthermore, these processes might even be more pronounced, that is, better identifiable during the second as during the first time-window as in the former no complex visual stimuli were presented. Consequently, we integrated time as a factor in our analyses as we expected effects of TPI in both time-windows or even an interaction between the two factors time-window and condition, as the effects of TPI might be differently pronounced in the two time-windows.

Greenhouse-Geyser corrections were performed on the *p*-values where necessary. For post-hoc pairwise comparisons (*t*-tests, two-tailed) of the ANOVA all *p*-values were Bonferroni-Holm corrected for multiple comparisons. Level of significance was set at $\alpha = .05$ for all analyses and partial eta-square ($\eta_p^2$) is reported as a measure of effect size for the ANOVAs.

Second, for exploratory data analyses and visualization, to inspect the localization of the theta ERS and the upper alpha ERD effects on the scalp, we calculated *topoplots* showing the ERD/ERS%-values of the theta and upper alpha frequency band for the two time-windows and three task conditions at all 27 electrode sites. Electrodes showing significantly different ERD/ERS%-values ($p < .05$) between task conditions (t-tests, two-sided, permutation-based statistics, using false-discovery rate to correct for multiple comparisons) are marked as red dots in the topoplots.

### 2.5.2. Fixation-related EEG data analyses

For the fixation-related EEG data analyses two rectangular areas of interest (AOIs) were defined, a picture-area and a text-area. The epoched and artifact-cleaned EEG data (resulting from the procedures described in 2.5.1) were re-epoched in two time-windows, each 1000 ms long (plus additional 250 ms before start and after end and 1000 ms of mirrored data each to avoid edge-artifacts), individually aligned to the first viewing of the text-AOI (i.e., first fixation on the text-AOI) and the second viewing of the picture-AOI (i.e., the first fixation on the picture-AOI after the text-AOI had been fixated at). Note that the typical viewing pattern of subjects started with fixations on the picture. This was because of the fixation-cross of the previous trials being situated within the picture-AOI. Then, subjects' fixations were on the text-AOI (indicating text reading), then back on the picture-AOI (see Table 1 for the average start-times of the initial fixations on the AOIs and Table 2 for the mean number of transitions between the two AOIs). For these two time-windows (each of 1000 ms length in the time-domain) the mean EEG frequency band power (Morlet wavelet convolution, width 8) and the ERD/ERS% data then were calculated. As baseline for calculating the ERD/ERS%-values served the averaged frequency band power of the two time-windows (i.e., the two viewings of the AOIs) and all three task conditions. The ERD/ERS% were calculated separately for each electrode and each frequency band. Note that the use of time-windows of 1000 ms length for the fixation-related analysis was a compromise between having enough data for calculating frequency band power data by Morlet wavelet convolution, while keeping the time-window adequately short to mostly capture mental processing demands for a certain AOI. We will elaborate more on that potential methodological limitation in Section 4.4.

Based on the recommendation of one anonymous reviewer, we additionally calculated the fixation-related ERD/ERS% by using the initial fixation-times on the AOIs as markers to select, for each trial separately, time-windows of 1000 ms length out of the TFRs reported above (see 2.5.1). In doing so, we avoided rerunning the Morlet wavelet decomposition for a second time and we were able to use the same baseline time-window for calculating the ERD/ERS% values as in the stimulus-locked EEG data analysis. However, as the outcomes of this alternative way of calculating the fixation-related ERD/ERS%-values did not substantially alter our initial results, we decided to mainly report the

**Table 1**
Mean start-times [ms] of the first and second viewing of the text and the picture.

| AOI | Match | | Partial match | | Mismatch | | N |
|---|---|---|---|---|---|---|---|
| 1st fixation picture | 0 | (0) | 0 | (0) | 0 | (0) | 25 |
| 1st fixation sentence | 630 | (141) | 650 | (168) | 622 | (188) | 25 |
| 2nd fixation picture | 1928 | (288) | 1940 | (313) | 1994 | (326) | 25 |
| 2nd fixation sentence | 2392 | (321) | 2482 | (339) | 2376 | (251) | 24 |

Note. The fixation-cross presented between the stimuli was positioned in the area of the picture, therefore for all participants the first fixations were on the picture AOI. One subject showed no detectable second fixations on the sentence, therefore for the second fixations only the data of 24 subjects is shown. Standard deviation is given in brackets.

[2] In addition, we calculated the TFRs using a different methodology, namely short-time fast-fourier transforms (FFT) with a sliding analysis-window (500 ms window-width, Hanning tapered, 20 ms time-steps). This approach resulted in virtually identical outcomes, underlining the robustness of our results. Yet for reasons of space we will only report the results of the Morlet wavelet convolution here. Note that in order to avoid edge artifacts we temporarily added a 2 s part of mirrored data each at the beginning and the end of each data epoch in the time-domain before conducting the wavelet decomposition (Cohen, 2014).

**Table 2**
Mean number of transitions between the picture-AOI and the text-AOI (*SD* is given in brackets).

| Match | | Partial match | | Mismatch | |
|---|---|---|---|---|---|
| 2.68 | (0.32) | 2.75 | (0.38) | 2.67 | (0.28) |

originally planned analyses and to report the results of the alternative analyses only briefly in addition.

As in the stimulus-locked analyses, for a comprehensive statistical analysis we conducted separate two-factorial repeated-measures ANOVAs for the theta and the upper alpha frequency band power at typical electrodes (Fz, Pz) with the factors *viewing* (first viewing of the sentence, second viewing of the picture) and *condition* (match, partial-match, mismatch). Additionally, as in the stimulus-locked analyses, we calculated topoplots to visualize the theta ERS and upper alpha ERD at the 27 electrode positions over the scalp.

## 3. Results

### 3.1. Manipulation check: sentence-verification test

The mean accuracies (i.e., the average percentage of correct responses for the sentences previously seen) in the sentence-verification test after the fifth learning block were 69.56% ($SD = 16.20$) for the stimuli of the match condition, 62.56% ($SD = 15.79$) for the stimuli of the partial-match condition, and 57.56% ($SD = 18.20$) for the stimuli of the mismatch condition. A one-factorial repeated-measures ANOVA revealed a main effect of condition, $F(2,48) = 7.04$, $p = .002$, $\eta_p^2 = .23$. Post-hoc pairwise comparisons (t-tests, two-sided) were Bonferroni-Holm corrected for multiple comparisons. Accuracies were significantly higher for the match condition as compared to the partial-match condition ($p = .040$) and the mismatch condition ($p = .008$), whereas the partial-match and mismatch condition did not differ ($p = .133$). This outcome is generally in line with literature (e.g., Arndt et al., 2015; Schüler et al., 2015), indicating that TPI becomes testable in the partial-match condition where the integrated memory representation of the textual and the pictorial information results in more false recognitions, that is, lower accuracies for this condition as compared to the match condition where TPI does not lead to such an error. The lower accuracies of the mismatch condition can be explained by the fact that two mental representations instead of one had to be memorized in this condition. Irrespective of the concrete task condition, the mean accuracies of the sentence-verification task of the current study are somewhat lower as the accuracies typically reported in literature (e.g., Arndt et al., 2015; Schüler et al., 2015). This difference might mainly be due to the fact that in the current study the sentences used as distractors were – unlike to the studies in literature – not completely new, previously unseen sentences, but sentences that had been presented in the blocks before, yet not in block 5 that was the block the participants were instructed to relate their decisions on the sentence-verification exclusively to. Another important result of the sentence-verification task was that the recall accuracies of the match and mismatch conditions were significantly above chance-level (all $p < .001$), providing a strong hint that the participants performed the learning task thoroughly as expected.

### 3.2. Physiological measures of mental processing demands

#### 3.2.1. EEG frequency band power: stimulus-locked analyses

The two-factorial repeated-measures ANOVAs that we conducted on the ERD/ERS% data revealed for the mean theta frequency band at electrode Fz (see Fig. 2, left-hand side) a main effect of time-window, $F(1,24) = 8.54$, $p = .007$, $\eta_p^2 = .26$. The theta ERS was more pronounced in the first time-window (i.e., during the stimulus presentation) as compared to the second time-window (i.e., during the fixation cross). In addition, we observed a main effect of condition, $F(2,48) = 5.75$, $p = .006$, $\eta_p^2 = .19$. Post-hoc pairwise comparisons of the main effect of condition indicated an increased theta ERS for the mismatch as compared to both, the partial-match condition ($p = .022$), and the match condition ($p = .030$). The match and partial-match condition were virtually identical ($p = .564$). The interaction between the two factors *time-window* and *condition* was not significant, $F < 1$, $p > .700$.

The two-factorial repeated-measure ANOVA that we conducted on the ERD/ERS% data for the upper alpha frequency band at Pz also revealed a main effect of time-window, $F(1, 24) = 87.83$, $p < .001$, $\eta_p^2 = .79$, and a main effect of condition, $F(2, 48) = 5.45$, $p = .007$, $\eta_p^2 = .19$. These main effects were qualified further by a significant interaction, $F(2, 48) = 4.69$, $p = .014$, $\eta_p^2 = .16$. Post-hoc pairwise comparisons showed that the upper alpha ERD generally was larger (i.e., more negative; see Fig. 2, right-hand side) during the first time-window (i.e., the stimulus presentation) as compared to the second time-window (i.e., the fixation cross; all $ps < .001$). During the first time-window, there was no difference between the three task conditions (all $p > .754$). In the second time-window, however, the upper alpha ERD was significantly more pronounced (as indicated by lower ERD/ERS%-values) in the mismatch condition as compared to both, the match condition ($p = .001$) and the partial-match condition ($p = .046$), with the latter two showing no significant difference ($p = .626$).

Although we did not observe a significant interaction in the ANOVA for the theta ERS as described above, the exploratory topoplots (see Fig. 3, left-hand side) indicated that the theta ERS was most prominent in the first time-window, especially at left-frontal but also right-parietal electrodes when comparing the mismatch and the match condition, respectively the partial-match condition. With respect to theta ERS there were only few electrodes showing a significant difference between the match condition and the partial-match condition, yet only in the second

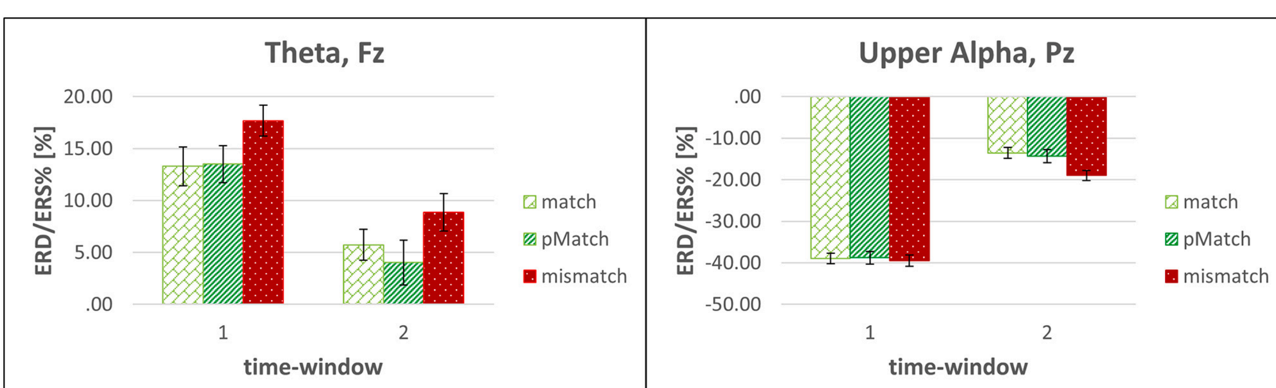


**Fig. 2.** Stimulus-locked EEG frequency band power analysis (ERD/ERS%-values) of the theta frequency band power at electrode Fz (left) and the upper alpha frequency band power at electrode Pz (right) for the first (0–5000 ms) and second (5000–6500 ms) time-window. Note. Error bars ±1 SEM.

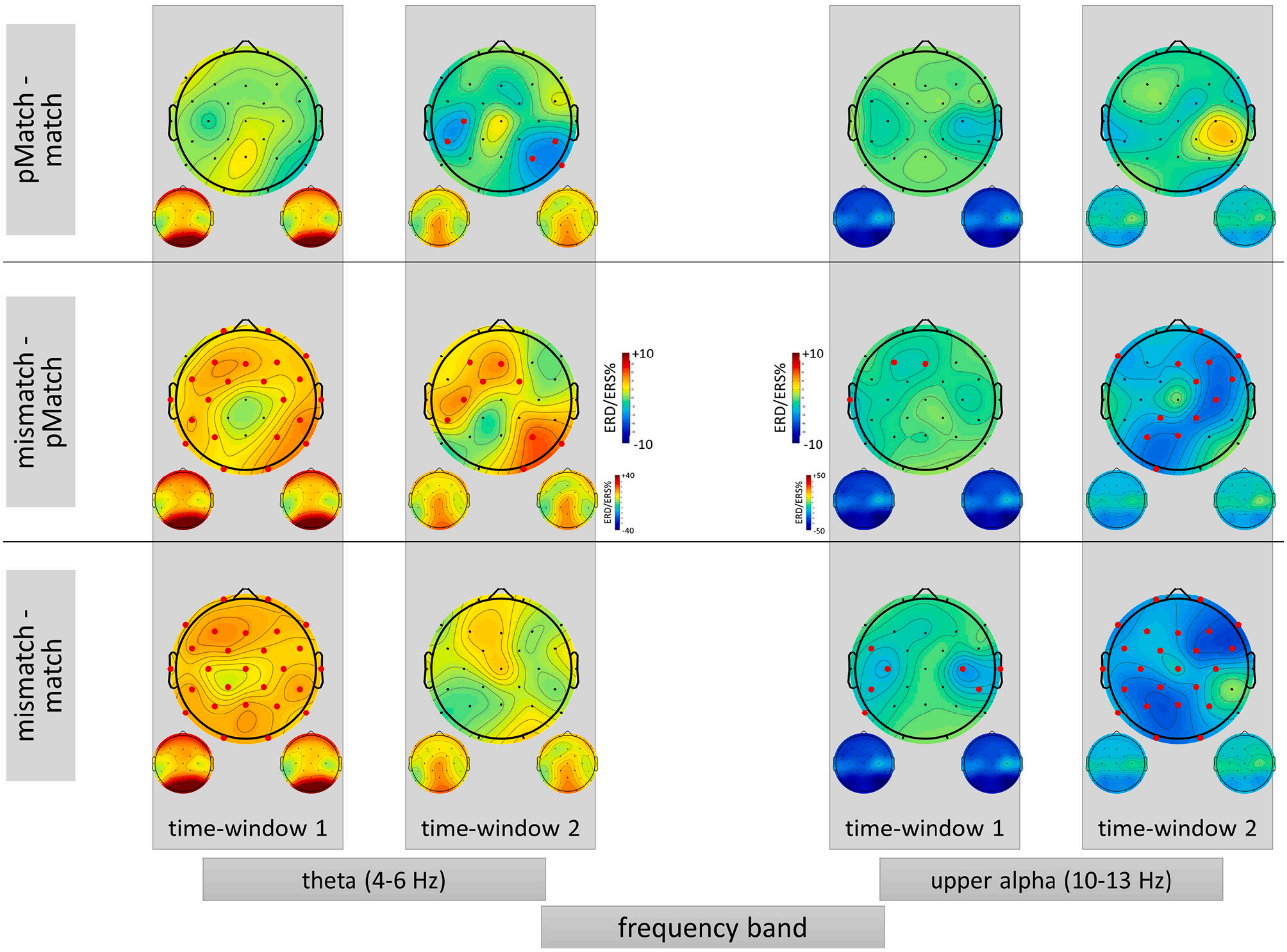


**Fig. 3.** Topoplots showing the ERD/ERS% values of all 27 electrode sites on the scalp for the theta (left part of the figure) and the upper alpha (right part of the figure) frequency band, the three match-conditions, and the two stimulus-locked time-windows. Blue color indicates ERD, red color ERS. The larger schematic heads show difference values of the ERD/ERS%-values between the two conditions indicated in the labels on the left-hand side of the figure (i.e., mismatch minus match for the lower row, mismatch minus partial-match for the middle row and partial-match minus match for the upper row). The smaller schematic heads on each side of a large difference-topoplot show the ERD/ERS%-values of the two corresponding conditions (i.e., mismatch and match, mismatch and partial-match, partial-match and match). The red dots in the larger schematic heads indicate electrodes showing significant different ERD/ERS%-values between adjacent task conditions (paired *t*-tests, two-sided, $p < .05$, using permutation-based statistics and false-discovery-rate correction for multiple comparisons). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

time-window.

For the upper alpha ERD the exploratory toploplots (see Fig. 3, right-hand side) show that the upper alpha ERD is most pronounced in the second time-window at mid- to left-parietal and right central and frontal electrodes. In the first time-window there are only for the comparison between match and mismatch some few electrodes indicating significant differences between the task conditions.

The time-frequency representations (TF-plots, see Fig. 4) also indicated significant differences in the theta band ERD/ERS%-values at electrode Fz between match and mismatch and partial-match and mismatch in a time-region at about 1000 ms to 2000 ms post-stimulus onset that was absent between match and partial-match. For the upper alpha ERD/ERS%-values at electrode Pz significant differences were seen between match and mismatch and partial-match and mismatch in a time-region of about 6000 ms to 6500 ms post-stimulus onset, a difference that again was absent between match and partial-match.

### 3.2.2. *EEG frequency band power: fixation-related analyses*

For the ERD/ERS%-values of the theta frequency band at the frontal electrode Fz (Fig. 5, upper half, left-hand side) a two-factorial repeated-measures ANOVA revealed a main effect of viewing, $F(1, 24) = 33.96$, $p < .001$, $\eta_p^2 = .59$, as well as a main effect of condition, $F(2, 48) = 8.03$, $p = .001$, $\eta_p^2 = .25$. The interaction between the two factors was not significant ($F < 1$, $p > .59$). The theta ERS was generally more pronounced during the first time-window (i.e., the first reading of the text-AOI) as compared to the second time-window (i.e., the second viewing of the picture). Interestingly, the theta ERS was more pronounced in the mismatch condition as compared to both, the match condition ($p = .004$) and the partial-match condition ($p = .003$), with the latter two showing no statistically significant differences ($p = .975$). This outcome is in line with the results of the stimulus-locked analysis.

The outcomes of the additional ANOVAs conducted on the alternatively calculated ERD/ERS%-values (see Fig. 5 lower part, left-hand side) were virtual identical, revealing a main effect of viewing, $F(1, 24) = 22.91$, $p < .001$, $\eta_p^2 = .52$, as well as a main effect of condition, $F(2, 48) = 9.12$, $p < .001$, $\eta_p^2 = .28$. The interaction between the two factors was not significant ($F < 1$, $p > .539$).

For the ERD/ERS%-values of the upper alpha frequency band at the parietal electrode Pz (Fig. 5, upper half, right-hand side) a two-factorial repeated-measures ANOVA revealed a main effect of viewing, $F(1, 24) = 6.58$, $p = .017$, $\eta_p^2 = .22$, with the upper alpha ERD generally being more pronounced in the second time-window (i.e. during the second viewing of the picture). The experimental manipulation did show no significant effect for the upper alpha frequency range, $F(2, 48) = 1.56$, $p$

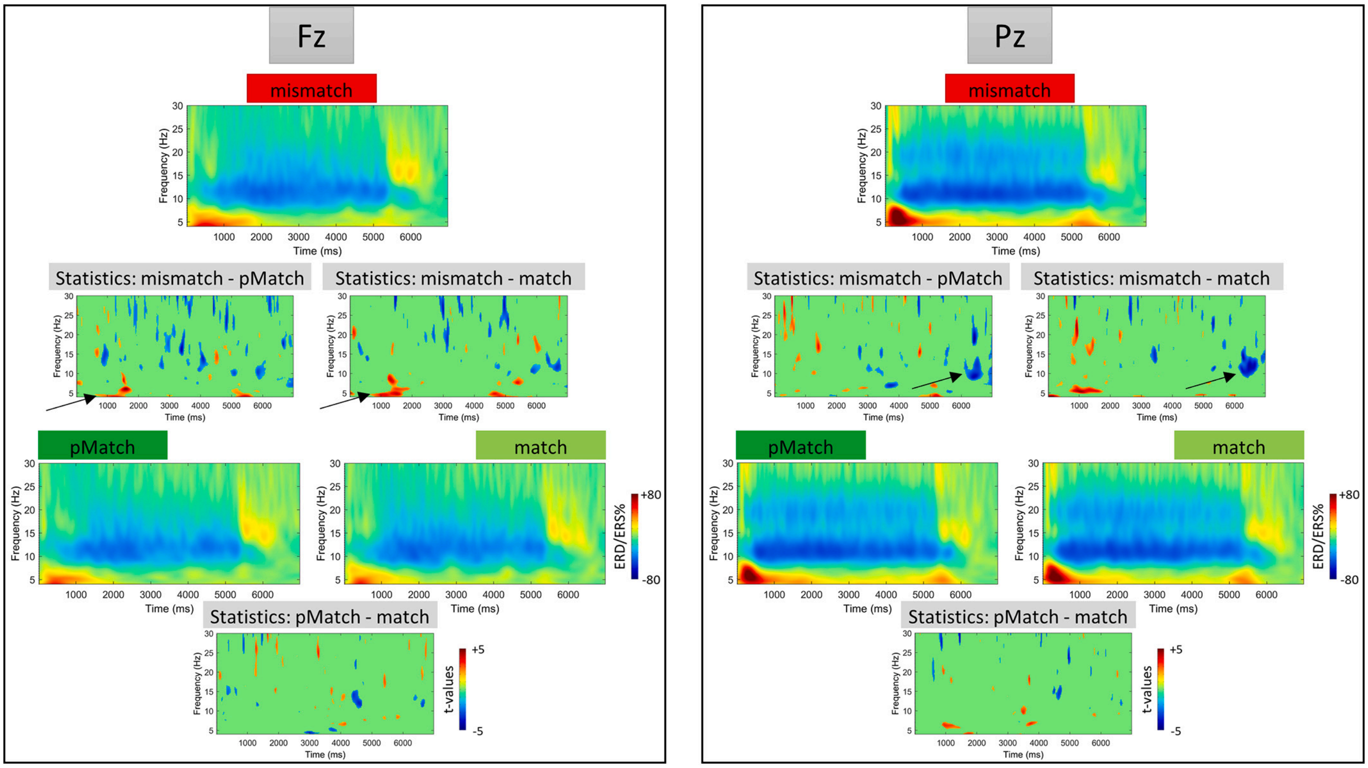


**Fig. 4.** Time-frequency representations (TF-plots) showing the ERD/ERS%-values at electrodes Fz (left-hand side) and Pz (right-hand side) over time (x-axis, 0–7 s post-stimulus onset) and the frequency range from 4 to 30 Hz (y-axis) of all three task conditions. Red color corresponds to positive ERD/ERS%-values, blue color to negative ERD/ERS%-values. The *statistics plots* show the *t*-values that were statistically significant ($p < .05$; paired t-tests, two-sided, using permutation-based statistics and false-discovery rate as correction for multiple comparisons) for the statistical comparison of the ERD/ERS%-values of adjacent task conditions (i.e., mismatch and partial-match, mismatch and match, partial-match and match). The small arrows in the *statistics plots* point to regions for which the ANOVAs on time-averaged data also showed significant effects. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

$= .222$, $\eta_p^2 = .06$. The interaction between the two factors also was not significant, $F(2, 48) = 1.21$, $p = .306$, $\eta_p^2 = .05$. This outcome is also in line with the stimulus-locked EEG data analysis.

The outcomes of the additional ANOVAs conducted on the alternatively calculated ERD/ERS%-values (see Fig. 5 lower part, right-hand side) showed neither a main effect of viewing, $F(1, 24) = 2.07$, $p = .163$, $\eta_p^2 = .08$, nor a main effect of condition, $F < 1$, $p > .591$, nor an interaction between the two factors, $F < 1$, $p > .396$. The missing effect for *viewing* in this analysis might be due to the use of a different baseline indicating the necessity to be cautious when interpreting the effect on *viewing* in the initial analysis reported above.

The exploratory topoplots (see Fig. 6, left-hand side) indicated the theta ERS to be most pronounced at (left-) frontal electrodes and also at some parietal electrodes for the comparison of mismatch and match and mismatch and partial-match conditions. On virtually no electrode there was a difference in theta ERS when comparing match and partial-match conditions.

Although the ANOVA reported above did not show a significant main effect of *condition* for the ERD/ERS%-values at electrode Pz in the upper alpha frequency range, especially during the first viewing of the sentence some mainly left-parietal electrodes showed a significant difference between the mismatch condition and the match and partial-match condition respectively. On these electrode locations there seemed to be an upper alpha ERS (rather than the ERD we expected) for the mismatch condition as compared to the match and partial-match conditions (see Fig. 6, right-hand side).

### 3.2.3. Eye-tracking data: pupil dilation

We calculated the average pupil dilation data for each trial stimulus-locked for two time-windows, namely the stimulus presentation and the following fixation cross (see Table 3). Note that in order to avoid potential spill-over effects of the previous trial on the pupil size due to the rather low reactivity of the pupil to mental processing demands, the range of the first time-window was set from 1000 to 5000 ms post-stimulus onset. A two-factorial repeated-measures ANOVA with the factor time-window (first, second) and condition (match, partial-match, mismatch) revealed a main effect of condition, $F(2,48) = 4.74$, $p = .013$, $\eta_p^2 = .17$. The pupil size was significantly larger in the mismatch condition as compared to the partial-match condition ($p = .011$), all other pairwise comparisons showing no significant difference ($p > .336$). There was no main effect for time-window ($F < 1.01$, $p > .33$) and no interaction ($F < 1.22$, $p > .30$) (Fig. 7).

## 4. Discussion

In the current study we used and combined eye-tracking measures and EEG measures to study TPI. To do so, we assessed the mental processing demands associated with three different types of text-picture combinations, namely combinations of textual and pictorial information that either matched (i.e., being identical), partially matched (i.e., being identical except for one complementary information in one of both representations), or mismatched (i.e., being disparate) using eye-tracking and EEG as process measures. As expected, we observed the mental processing demands to be highest in the mismatch condition, when TPI was not possible and two completely disparate mental representations (i.e., the pictorial and the textual mental representation) had to be mentally constructed and memorized. This pattern was confirmed by all of the process measures that were related to mental processing demands that we analyzed (for a detailed discussion see below). Additionally, we did not observe a significant difference


**Fig. 5.** Fixation-related EEG frequency band power analysis showing ERD/ERS%-values of the theta frequency band at electrode Fz (left) and the upper alpha frequency band at electrode Pz (right) for the first viewing of the sentence and the second viewing of the picture. The upper half of the figure shows the fixation-related ERD/ERS%-values calculated as described in Section 2.5.2 with an averaged whole-epoch baseline, the lower half of the figure shows the fixation-related ERD/ERS%-values alternatively calculated based on the recommendation of one anonymous reviewer using the baseline from the stimulus-locked analysis. Irrespectively of the baseline time-period and method used for calculating the fixation-related ERD/ERS%-values, the most important outcomes with regard to our experimental manipulation were virtually not different. Note. Error bars ±1 SEM.

between the match condition and the partial-match condition. This finding indicates that integration occurred in both conditions to a similar degree, regardless of whether text and picture matched or only partially matched. Hence, adding an additional piece of information to one of the two representations did not result in measurable increased mental processing demands. This indicates that fine-grained differences between text-picture overlap do not change the mental processing demands.

In both, the stimulus-locked analysis as well as the fixation-related analysis, the EEG frequency band power proved to be sensitive for assessing the increased mental processing demands in the mismatch condition as compared to the two other conditions. These outcomes substantiated the eye-tracking results (i.e., the pupil dilation data), as, noteworthy, the EEG frequency band power measures are clearly more directly related to underlying brain activity as compared to the eye-tracking measures. In turn, the outcomes also showed the feasibility and validity of using EEG frequency band power measures in the study of rather complex task materials.

### 4.1. Stimulus-locked EEG data analysis

The classical, stimulus-locked EEG data analysis revealed an overall increased theta ERS at frontal electrodes for the mismatch (i.e., TPI not possible) as compared to the partial-match condition and, yet only numerically, to the match condition (i.e., the conditions where TPI was possible), with the latter two showing virtually no differences. This outcome of increased processing demands in the mismatch condition as compared to match and partial-match conditions was corroborated further by the upper alpha ERD at parietal electrodes that was most pronounced in the mismatch as compared to both, the match and partial-match condition. Important to note, the upper alpha frequency band power did only show a significant decrease (i.e., a more pronounced ERD) for the mismatch condition as compared to the other two conditions in the second time-window, when only a fixation-cross was presented. In the first time-window, when the stimuli were presented, the upper alpha frequency band power did not show any differences between the task conditions.

This outcome is interesting as it indicates a functional difference between theta ERS and (upper) alpha ERD with respect to potentially underlying cognitive and hence neuronal mechanisms. Oscillatory activity in the theta band has been especially associated with processes of working memory and cognitive control (e.g., Sauseng et al., 2010, 2005). Thus, the increased theta ERS early in time in the mismatch condition might be due to the detection of the non-corresponding information conveyed through text and picture, that is, the impossibility for TPI to occur and the subsequently higher working memory load as two different representations have to be processed and memorized further. Especially, the theta ERS in the mismatch condition might be indicative for subjects' detection and processing of non-corresponding information (Cohen, 2011; Luu et al., 2004; Nigbur et al., 2011). Instead, while also indicative for mental processing demands, upper alpha power might specifically reflect semantic memory demands (Klimesch et al., 1997a; Klimesch, 1999). Thus, the increased upper alpha ERD in the mismatch condition as compared to the two other conditions that was visible only during the second time-window, might reflect the interactions between working and long-term memory representations, and the higher mental processing demands when two semantically different mental representations have to be memorized separately.

Another, yet similar interpretation of the observed result pattern of the theta and upper alpha frequency band power effects might be that

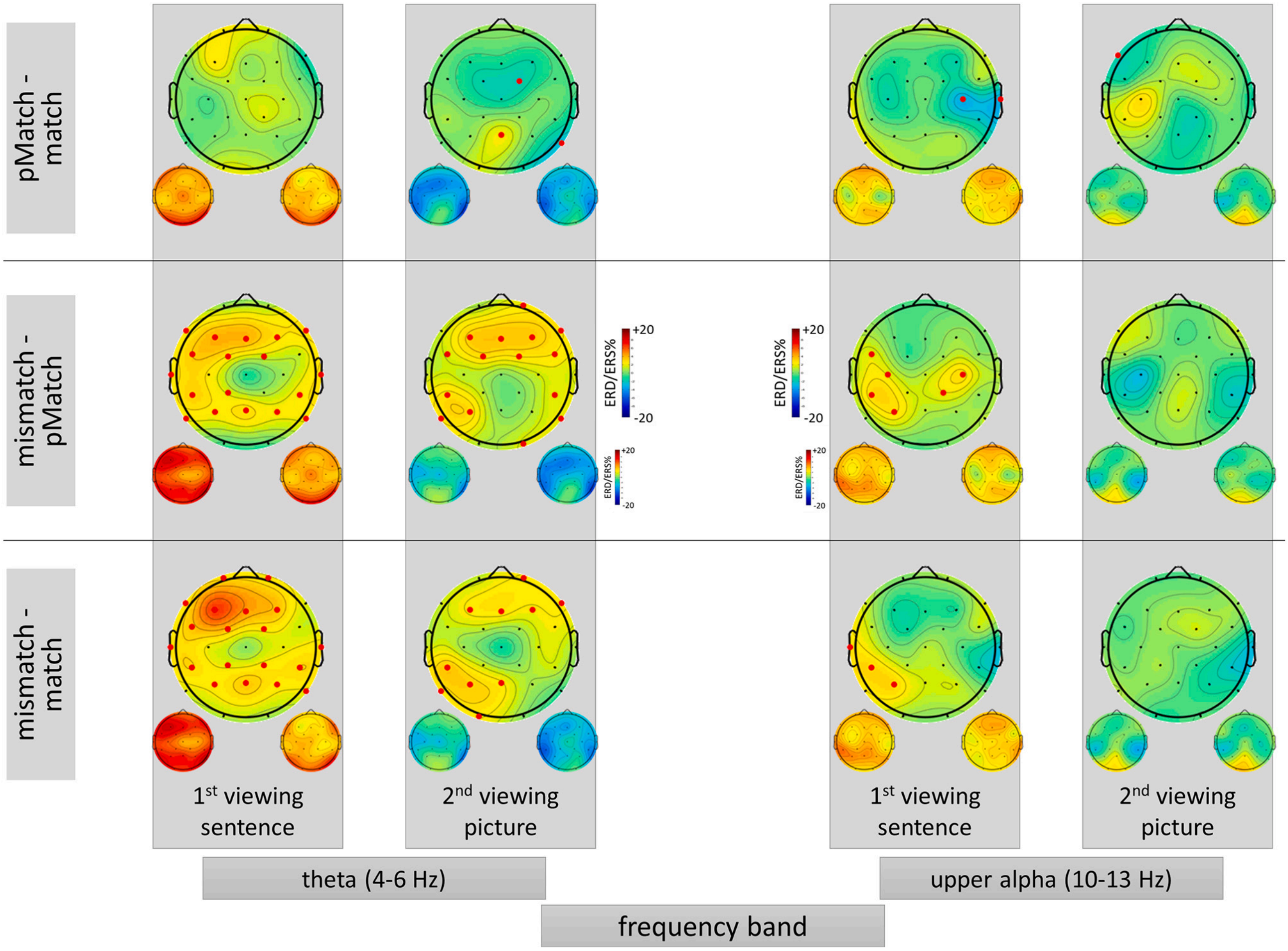


**Fig. 6.** Topoplots showing the ERD/ERS% values of all 27 electrode sites on the scalp for the theta (left part of the figure) and the upper alpha (right part of the figure) frequency band for the first viewing of the sentence and the second viewing of the picture for the three task conditions. Blue color indicates ERD, red color ERS. The larger schematic heads show difference values of the ERD/ERS%-values between the two conditions indicated in the labels on the left-hand side of the figure (i.e., mismatch minus match for the lower row, mismatch minus partial-match for the middle row and partial-match minus match for the upper row). The smaller schematic heads on each side of a large difference-topoplot show the ERD/ERS%-values of the two corresponding conditions (i.e., mismatch and match, mismatch and partial-match, partial-match and match). The red dots in the larger topoplots indicate electrodes that show significant different ERD/ERS%-values between adjacent task conditions (paired *t*-tests, two-sided, $p < .05$, using permutation-based statistics and false-discovery-rate correction for multiple comparisons). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

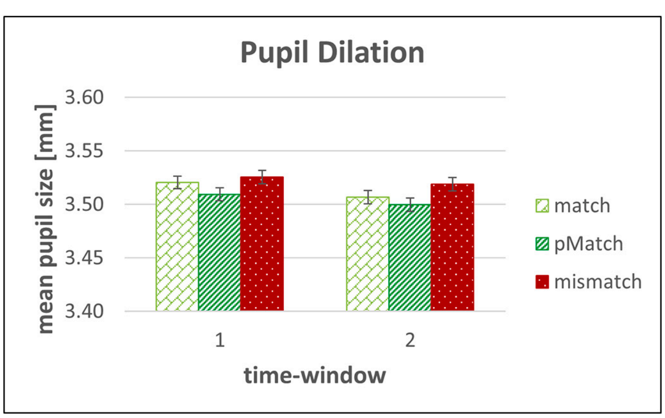


**Fig. 7.** Mean pupil size [mm] for the three task conditions during stimulus presentation (time-window 1) and the following fixation cross (time-window 2). Note. Error bars ±1 SEM.

while oscillatory activity in the theta frequency band might indicate early cognitive processing like the detection of incongruent information, the oscillatory activity in the upper alpha frequency band range might reflect later cognitive processes like semantic integration in memory. The mismatch condition might take some more processing time because of the incongruency between text and picture, thus resulting in an upper alpha effect that is only visible in the second time-window (i.e., during the post-stimulus fixation-cross).

An additional observation has to be discussed, namely the general difference in the upper alpha ERD (and theta ERS) between the first and second time-window. Overall, the upper alpha ERD and the theta ERS is more pronounced in the first time-window as compared to the second time-window. This might be due to the fact that during the first time-window the stimulus materials were presented (i.e., the text-picture combinations) whereas during the second time-window only a fixation-cross was shown (i.e., almost no visual input). Consequently, the second time-window might be considered as a purely retention phase where subjects tried to memorize the pictorial and textual information they had processed before. Interestingly, the observed reduced ERD in the upper alpha frequency band in the second time-window as compared to the first time-window can be seen to be in line with literature showing even an alpha ERS during the retention phase of working memory tasks (Jensen et al., 2002; Klimesch et al., 2007; Palva and Palva, 2007).

### 4.2. Fixation-related EEG frequency band power analysis

In line with the stimulus-locked analysis, the fixation-related analysis revealed an increased theta ERS at frontal electrodes for the mismatch condition as compared to both, the partial-match and the match condition. This effect occurred with the first reading of the textual information and persisted during the following viewing of the picture. Note that we use the term theta ERS, although numerically in the second time-window of the initial analysis all ERD/ERS%-values are negative due to the averaged whole-epoch baseline (see Fig. 5, upper part, left-hand side). Nevertheless, in the mismatch condition the ERD/ERS %-values of the theta frequency band are less negative as compared to both, match and partial-match, in our view justifying the use of the term theta ERS here, too (i.e., a relative theta ERS). This interpretation is corroborated by the additional alternatively calculated fixation-related ERD/ERS%-values (see Fig. 5, lower part, left-hand side). Here for the viewing of both AOIs a theta ERS clearly is visible.

As in the stimulus-locked analysis for the first time-window (i.e., the stimulus presentation), the upper alpha frequency band power did not show a significant difference between the task conditions at the parietal electrode Pz in the overall fixation-related analysis either. Yet, the exploratory topoplots indicated at left-parietal electrode locations an increased upper alpha ERS for the mismatch condition as compared to the match and partial-match condition during the first reading of the text, as well as an increased upper alpha ERS at fronto-parietal electrode locations during the second viewing of the picture. This outcome might be seen in contrast to our hypothesized ERD-effect for increased mental processing demands when TPI is not possible. However, as alpha ERS has been described to indicate internal inhibitory processes (Klimesch et al., 2007; Klimesch, 2012; Palva and Palva, 2007), the observed upper alpha ERS might be related to internal mental processes of keeping pictorial and textual information separate entities, that is, the inhibition of one information channel (e.g., the pictorial information) overriding the other one (e.g., the textual information). Clearly, such an interpretation is highly speculative based on our data and would have to be systematically studied further in order to understand this effect.

In sum, the fixation-related analysis closely resembled the stimulus-locked analysis, indicating the validity of the fixation-related analysis. This outcome is noteworthy, as in more ecological valid multimedia task settings (as compared to the comparably highly controlled task setting we used in the current study), a classical, stimulus-locked EEG data analysis might not be feasible. This is because, for example, the "stimulus" (i.e., the multimedia materials) might consist of a longer text and several pictures that are presented simultaneously on the screen while allowing participants a self-paced reading. For these complex, more realistic task materials a fixation-related EEG frequency band power analysis might be a promising methodology to assess subjects' mental processing demands in relation to the different parts of the multimedia materials that are fixated at (for a more detailed discussion of the potentials of fixation-related EEG frequency band power analysis see Scharinger, 2018).

At this point one might wonder why the fixation-related data analyses showed virtually no different outcomes as compared to the stimulus-onset locked analyses. On the contrary, one even might have expected different outcome as the fixation-related analyses is closely connected to certain fixations on the text and on the picture and thus specific cognitive processes might show up in the EEG data that might differ for example between text reading and picture viewing. In our opinion, in contrast to fixation-related potentials (e.g., Dimigen et al., 2011), EEG frequency band power data provides a rather rough picture of cognitive processes in time. Calculating TFRs (i.e., the power spectrum of frequencies over time) requires for technical reasons a certain amount of data of the time domain, that is, the time resolution of TFRs (i.e., in the frequency domain) is not as good as the time-resolution in the time domain. We used a 1000 ms time-window as input to calculate the fixation-related frequency band power. Thus, the fixation-related frequency band power analyses covered already one fifth of the data window used in the stimulus-onset based data analyses (with respect to the 5000 ms of the stimulus presentation). This might be a central reason why we did observe the same pattern of results in both data analyses. Importantly however, as described in the introduction we are quite convinced that EEG frequency band power per se is a valid and promising measure of mental processing demands. Fixation-related analyses of EEG frequency band power might be especially of value in completely unconstrained, free viewing situations of complex task materials. In the current task paradigm, with a rather short and fixed presentation time of stimuli the fixation-related EEG frequency band power analyses might be of limited additional value. Importantly however, the current study demonstrated the general feasibility and validity of such an analysis that might become especially of interest for studying complex task materials.

### 4.3. Pupil dilation

As already stated above, the outcomes of the eye-tracking measures were mostly in line with the EEG results. For the pupil dilation we observed an overall effect of task condition on pupil size, with the pupil dilation being more pronounced in the mismatch condition as compared to the partial-match condition. Yet, there was no statistically significant difference between the mismatch condition and the match condition, indicating that for the current task materials and task settings the pupil dilation might not have been sensitive enough to detect differences between all the task conditions. In comparison, the EEG frequency band power data in the current study seems to be more sensitive. In turn, this outcome can be considered to underline the added value of using more directly brain-related measures of mental processing demands like the EEG theta and (upper) alpha frequency band power when studying multimedia task materials.

### 4.4. Limitations of the current study

Some limitations of the current study have to be addressed. First, as described in the Method section, for reasons of comparability with previous studies (Arndt et al., 2015; Schüler et al., 2015, 2019) we had only a limited number of different topics depicted in the stimuli. Therefore, over the course of the five learning blocks, the same textual or pictorial information could occur for several times, yet each time in another condition. Consequently, from block two on, most of the stimuli participants had to learn had previously been encountered, yet in another combination. Thus, starting with block two, participants not only had to learn and integrate textual and pictorial information but potentially also to override old information in memory (i.e., previous text-picture combinations of the same topics). Therefore, in the current study the learning process might have been more complex as compared to a situation of only learning completely new stimuli. The repetition of stimuli-parts (e.g., the same picture combined with different sentences forming different task conditions) during the learning phase and potential interference effects thereof clearly are a central limitation of the current study. For identifying and disentangling pure processes of TPI and potential interference effects of such a partly repetition of textual or pictorial information, future studies should be conducted that might compare two learning situations: One with and one without the repetition of stimuli-parts. Nevertheless, although we cannot disentangle whether our observed effects are mainly due to pure processes of TPI or mainly due to processes of TPI that include the (partial) mental replacement of previously memorized information, we think the current study is of value. This is because the current study adds direct process measures like the EEG and especially the EEG frequency band power data to research in instructional psychology on TPI (Arndt et al., 2015; Schüler et al., 2015, 2019), underlining the results reported there.

Furthermore, the learning process we created might be ecologically more valid as learning completely from the scratch seldom occurs for adults. More importantly however, we are quite convinced that the interpretation of our data with respect to mental processing demands and TPI is valid irrespective whether potential processes of information overriding in long-term memory took place or not. This is because by task design, all three task conditions were equally distributed with respect to the used and in parts repeated text-picture combinations.

Second, although (as stated above) we have created a typical learning situation where new information can be in conflict with previously memorized information leading to the necessity to update previously stored information, with respect to the task materials the ecological validity of the current study might be seen critically. This is because instead of 'true' multimedia materials, we used very reduced stimuli, consisting of only simple, black-and-white line-drawings and few textual information. In addition, the artificial task presentation of a high number of stimuli with a rather short, fixed presentation time might be considered critically. However, in line with recommendations in the literature (e.g., Gerjets et al., 2014), we strongly belief that in order to be able to meaningfully interpret EEG data it is important to record valid and largely enough data sets for analyses, avoiding potentially confounds as far as possible, that is, using controlled, yet potentially rather artificially task materials and task designs. Especially, this might be important for initial studies using new methodologies (like, in this study, fixation-related EEG frequency band analysis). Future studies then might use the methodology to study TPI in more realistic multimedia task materials.

Third, we would like to point out a general problematic of the fixation-related EEG data analysis one has to be aware of. For a reliable calculation of the EEG frequency band power with an adequate frequency resolution a certain length of the time-window used is required. As described in the Method section, we therefore decided to use time-windows of 1000 ms length. Typical fixation durations for text reading or picture viewing are between 250 and 300 ms (Rayner, 2009). Therefore, within the time-window of 1000 ms used for the EEG data analysis time-locked with a certain fixation on a specific AOI (e.g., the text) there might not only be fixations on this specific AOI but also on other locations on the screen (e.g., the picture). In other words, the fixation-related EEG frequency band power analysis provides a rather rough picture of the mental processing load related to the viewing of specific AOIs. As indicated by Table 1 for the first time-window of analysis, the first viewing of the text, we might have predominantly captured mental processing demands for reading after some initial fixations on the picture. This is because, on average the fixations on the text sequentially summed up to more than 1000 ms in duration. Yet, the second time-window of analysis, the second viewing of the picture, with mean viewing times of about 600 ms, clearly entails other fixations than those on the picture only (and hence further cognitive processing). Despite this clear limitation of the fixation-related frequency band power analysis, we nevertheless are quite convinced that this methodology can provide more specific insights into the current mental processing demands linked to the visual processing of different elements on the screen, thus enriching a classical stimulus-locked analysis. Particularly, a classical stimulus-locked data analysis might not always be possible in ecologically more valid task settings of self-paced multimedia learning. Then, the fixation-related frequency band power analysis might be a valuable methodology to study the mental processing demands during learning (which in turn is important to analyze and improve the design of multimedia learning materials).

## 5. Conclusion

In the current study we were interested in assessing the mental processing demands during TPI using eye-tracking and EEG frequency band power as process measures. We observed on virtually all measures we analyzed increased mental processing demands for the mismatch condition (i.e., when textual and pictorial information was disparate and TPI was not possible) as compared to the match condition (i.e., textual and pictorial information being identical, TPI being possible) and the partial-match condition (i.e., textual and pictorial information being identical except for one complementary information in one of the representations, TPI being possible). The outcomes of the current study thus underline the assumptions that – if the materials allow for TPI - TPI takes place, resulting in overall reduced mental processing demands as compared to text-picture combinations where TPI is not possible. Furthermore, the current study showed the validity and feasibility of using EEG frequency band power to assess mental processing demands in complex multimedia task materials, corroborating and extending pure eye-tracking process measures. Finally, the fixation-related EEG data analysis proved to be a valuable account for studying multimedia learning (like text-picture combinations) in free viewing situations.

## Compliance with ethical standards

Funding: This study was funded by the Leibniz ScienceCampus Tübingen "Informational Environments". The funding source had no involvement in the study design, in the collection, analysis and interpretation of data, in the writing of the report, nor in the decision to submit the article for publication.

Conflict of interest: All authors declare to have no conflict of interest.

Ethical approval: All procedures performed in the study involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed consent: Informed consent was obtained from all individual participants included in the study.

## References

Anglin, G.J., Vaez, H., Cunningham, K.L., 2004. Visual representations and learning: The role of static and animated graphics. In: Handbook of Research on Educational Communications and Technology, 2nd edn. Erlbaum, Mahwah, NJ, pp. 865–916.

Antonenko, P., Niederhauser, D.S., 2010. The influence of leads on cognitive load and learning in a hypertext environment. Comput. Hum. Behav. 26 (2), 140–150. https://doi.org/10.1016/j.chb.2009.10.014.

Antonenko, P., Paas, F., Grabner, R., van Gog, T., 2010. Using electroencephalography to measure cognitive load. Educ. Psychol. Rev. 22 (4), 425–438. https://doi.org/10.1007/s10648-010-9130-y.

Arndt, J., Schüler, A., Scheiter, K., 2015. Text-picture integration: how delayed testing moderates recognition of pictorial information in multimedia learning. Appl. Cogn. Psychol. 29 (5), 702–712. https://doi.org/10.1002/acp.3154.

Baddeley, A., 2012. Working memory: theories, models, and controversies. Annu. Rev. Psychol. 63, 1–29. https://doi.org/10.1146/annurev-psych-120710-100422.

Baddeley, A., Hitch, G., 1974. Working memory. Psychol. Learn. Motiv. 8, 47–89. https://doi.org/10.1016/S0079-7421(08)60452-1.

Brouwer, A.M., Hogervorst, M.A., Holewijn, M., van Erp, J.B.F., 2014. Evidence for effects of task difficulty but not learning on neurophysiological variables associated with effort. Int. J. Psychophysiol. 93 (2), 242–252. https://doi.org/10.1016/j.ijpsycho.2014.05.004.

Butcher, K.R., 2014. The multimedia principle. In: Mayer, R. (Ed.), The Cambridge Handbook of Multimedia Learning. Cambridge University Press, Cambridge, pp. 174–205. https://doi.org/10.1017/CBO9781139547369.010.

Cabestrero, R., Crespo, A., Quirós, P., 2009. Pupillary dilation as an index of task demands. Percept. Mot. Skills 109, 664–678. https://doi.org/10.2466/pms.109.3.664-678.

Cavanagh, J.F., Frank, M.J., 2014. Frontal theta as a mechanism for cognitive control. Trends Cogn. Sci. 18 (8), 414–421. https://doi.org/10.1016/j.tics.2014.04.012.

Cohen, M.X., 2011. Error-related medial frontal theta activity predicts cingulate-related structural connectivity. NeuroImage 55 (3), 1373–1383. https://doi.org/10.1016/j.neuroimage.2010.12.072.

Cohen, M.X., 2014. Analyzing Neural Time Series Data. Theory and Practice. The MIT Press, Cambridge, Massachusetts.

Coppens, L.C., Gootjes, L., Zwaan, R.a., 2012. Incidental picture exposure affects later reading: evidence from the N400. Brain Lang. 122 (1), 64–69. https://doi.org/10.1016/j.bandl.2012.04.006.

Delorme, A., Makeig, S., 2004. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J. Neurosci. Methods 134 (1), 9–21. https://doi.org/10.1016/j.jneumeth.2003.10.009.

Delorme, A., Sejnowski, T., Makeig, S., 2007. Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis. NeuroImage 34 (4), 1443–1449. https://doi.org/10.1016/j.neuroimage.2006.11.004.

Dimigen, O., Sommer, W., Hohlfeld, A., Jacobs, A.M., Kliegl, R., 2011. Coregistration of eye movements and EEG in natural reading: analyses and review. J. Exp. Psychol. Gen. 140 (4), 552–572. https://doi.org/10.1037/a0023885.
Duncan, C.C., Barry, R.J., Connolly, J.F., Fischer, C., Michie, P.T., Näätänen, R., Polich, J., Reinvang, I., Van Petten, C., 2009. Event-related potentials in clinical research: guidelines for eliciting, recording, and quantifying mismatch negativity, P300, and N400. Clin. Neurophysiol. 120 (11), 1883–1908. https://doi.org/10.1016/j.clinph.2009.07.045.
Fairclough, S.H., Ewing, K., 2016. The effect of task demand and incentive on neurophysiological and cardiovascular markers of effort. Int. J. Psychophysiol. 1–9. https://doi.org/10.1016/j.ijpsycho.2017.01.007.
Gerjets, P., Walter, C., Rosenstiel, W., Bogdan, M., Zander, T.O., 2014. 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). https://doi.org/10.3389/fnins.2014.00385.
Gevins, A., Smith, M.E., 2000. Neurophysiological measures of working memory and individual differences in cognitive ability and cognitive style. Cereb. Cortex 10 (9), 829–839. https://doi.org/10.1093/cercor/10.9.829.
Harteis, C., Kok, E., Jarodzka, H., 2018. Editorial the journey to proficiency: exploring new objective methodologies to capture the process of learning and professional development. Frontline Learn. Res. 6 (3), 1–5. https://doi.org/10.14786/flr.v6i3.435.
Hyönä, J., 2010. The use of eye movements in the study of multimedia learning. Learn. Instr. 20 (2), 172–176. https://doi.org/10.1016/j.learninstruc.2009.02.013.
Jasper, H.H., 1958. The ten-twenty electrode system of the international federation. Electroencephalogr. Clin. Neurophysiol. 10, 371–375.
Jensen, O., Gelfand, J., Kounios, J., Lisman, J.E., 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. https://doi.org/10.1093/cercor/12.8.877.
Jung, T.P., Makeig, S., Humphries, C., Lee, T.W., Mckeown, M.J., Iragui, V., Sejnowski, T. J., 2000. Removing electroencephalographic artifacts by blind source separation. Psychophysiology 37 (2), 163–178. https://doi.org/10.1017/S0048577200980259.
Just, M.A., Carpenter, P.A., 1980. A theory of reading: from eye fixations to comprehension. Psychol. Rev. 87 (4), 329–354. https://doi.org/10.1037/0033-295X.87.4.329.
Kahneman, D., Beatty, J., 1966. Pupil diameter and load on memory. Science 154 (3756), 1583–1585. https://doi.org/10.1126/science.154.3756.1583.
Klimesch, W. (1999). EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. Brain Res. Rev., 29(2–3), 169–195. https://doi.org/10.1016/S0165-0173(98)00056-3.
Klimesch, W., 2012. Alpha-band oscillations, attention, and controlled access to stored information. Trends Cogn. Sci. 16 (12), 606–617. https://doi.org/10.1016/j.tics.2012.10.007.
Klimesch, W., Doppelmayr, M., Pachinger, T., Ripper, B., 1997a. Brain oscillations and human memory: EEG correlates in the upper alpha and theta band. Neurosci. Lett. 238 (1–2), 9–12. https://doi.org/10.1016/S0304-3940(97)00771-4.
Klimesch, W., Doppelmayr, M., Pachinger, T., Russegger, H., 1997b. Event-related desynchronization in the alpha band and the processing of semantic information. Brain Res. Cogn. Brain Res. 6 (2), 83–94. https://doi.org/10.1016/S0926-6410(97)00018-9.
Klimesch, W., Sauseng, P., Hanslmayr, S., 2007. EEG alpha oscillations: the inhibition–timing hypothesis. Brain Res. Rev. 53, 63–88. https://doi.org/10.1016/j.brainresrev.2006.06.003.
Knoeferle, P., Urbach, T.P., Kutas, M., 2011. Comprehending how visual context influences incremental sentence processing: insights from ERPs and picture-sentence verification. Psychophysiology 48 (4), 495–506. https://doi.org/10.1111/j.1469-8986.2010.01080.x.
Krause, C.M., 2003. Brain electric oscillations and cognitive processes. In: Hugdahl, K. (Ed.), Neuropsychology and Cognition. Experimental Methods in Neuropsychology, 21st ed. Kluwer Academic Publishers Group, Boston, MA, pp. 111–130.
Laeng, B., Sirois, S., Gredeback, G., 2012. Pupillometry: a window to the preconscious? Perspect. Psychol. Sci. 7 (1), 18–27. https://doi.org/10.1177/1745691611427305.
Li, S., Chen, S., Zhang, H., Zhao, Q., Zhou, Z., Huang, F., Hong, J., 2020. Dynamic cognitive processes of text-picture integration revealed by event-related potentials. Brain Research 1726 (April 2019), 146513. https://doi.org/10.1016/j.brainres.2019.146513.
Luu, P., Tucker, D.M., Makeig, S., 2004. Frontal midline theta and the error-related negativity: neurophysiological mechanisms of action regulation. Clin. Neurophysiol. 115 (8), 1821–1835. https://doi.org/10.1016/j.clinph.2004.03.031.
Makransky, G., Terkildsen, T.S., Mayer, R.E., 2019. Role of subjective and objective measures of cognitive processing during learning in explaining the spatial contiguity effect. Learn. Instr. 61 (January), 23–34. https://doi.org/10.1016/j.learninstruc.2018.12.001.
Mathôt, S., 2018. Pupillometry: psychology, physiology, and function. J. Cogn. 1 (1), 1–23. https://doi.org/10.5334/joc.18.
Mayer, R.E., 2009. Multimedia Learning, 2nd ed. Cambridge University Press, New York, NY.
Nigbur, R., Ivanova, G., Stürmer, B., 2011. Theta power as a marker for cognitive interference. Clin. Neurophysiol. 122 (11), 2185–2194. https://doi.org/10.1016/j.clinph.2011.03.030.
Oldfield, R., 1971. The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia 9, 97–113.
Palomäki, J., Kivikangas, M., Alafuzoff, A., Hakala, T., Krause, C.M., 2012. Brain oscillatory 4-35 Hz EEG responses during an n-back task with complex visual stimuli. Neurosci. Lett. 516 (1), 141–145. https://doi.org/10.1016/j.neulet.2012.03.076.
Palva, S., Palva, J.M., 2007. New vistas for alpha-frequency band oscillations. Trends Neurosci. 30 (4), 150–158. https://doi.org/10.1016/j.tins.2007.02.001.
Pesonen, M., Hämäläinen, H., Krause, C.M., 2007. Brain oscillatory 4-30 Hz responses during a visual n-back memory task with varying memory load. Brain Res. 1138, 171–177. https://doi.org/10.1016/j.brainres.2006.12.076.
Pfurtscheller, G., Lopes da Silva, F.H., 1999. Event-related EEG/MEG synchronization and desynchronization: basic principles. Clin. Neurophysiol. 110 (11), 1842–1857. https://doi.org/10.1016/S1388-2457(99)00141-8.
Rayner, K., 2009. Eye movements and attention in reading, scene perception, and visual search. Q. J. Exp. Psychol. 62 (8), 1457–1506. https://doi.org/10.1080/17470210902816461.
Sauseng, P., Klimesch, W., Schabus, M., Doppelmayr, M., 2005. Fronto-parietal EEG coherence in theta and upper alpha reflect central executive functions of working memory. Int. J. Psychophysiol. 57 (2), 97–103. https://doi.org/10.1016/j.ijpsycho.2005.03.018.
Sauseng, P., Griesmayr, B., Freunberger, R., Klimesch, W., 2010. Control mechanisms in working memory: a possible function of EEG theta oscillations. Neurosci. Biobehav. Rev. 34 (7), 1015–1022. https://doi.org/10.1016/j.neubiorev.2009.12.006.
Scharinger, C., 2018. Fixation-related EEG frequency band power analysis: a promising methodology for studying instructional design effects of multimedia learning material. Frontline Learn. Res. 6 (3), 57–71. https://doi.org/10.14786/flr.v6i3.373.
Scharinger, C., Kammerer, Y., Gerjets, P., 2015a. 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. https://doi.org/10.1371/journal.pone.0130608.
Scharinger, C., Soutschek, A., Schubert, T., Gerjets, P., 2015b. When flanker meets the n-back: what EEG and pupil dilation data reveal about the interplay between the two central-executive working memory functions inhibition and updating. Psychophysiology 52 (10), 1293–1304. https://doi.org/10.1111/psyp.12500.
Schnotz, W., 2014. An integrated model of text and picture comprehension. In: The Cambridge Handbook of Multimedia Learning, 2nd ed., pp. 49–69 Cambridge. https://doi.org/10.1017/CBO9780511816819.005.
Schüler, A., Arndt, J., Scheiter, K., 2015. Processing multimedia material: does integration of text and pictures result in a single or two interconnected mental representations? Learn. Instr. 35, 62–72. https://doi.org/10.1016/j.learninstruc.2014.09.005.
Schüler, A., Arndt, J., Scheiter, K., 2019. Does text–picture integration also occur with longer text segments? Appl. Cogn. Psychol. 33 (6), 1137–1146. https://doi.org/10.1002/acp.3558.
Siegle, G.J., Steinhauer, S.R., Stenger, V.A., Konecky, R., Carter, C.S., 2003. Use of concurrent pupil dilation assessment to inform interpretation and analysis of fMRI data. NeuroImage 20 (1), 114–124. https://doi.org/10.1016/S1053-8119(03)00298-2.
van der Wel, P., van Steenbergen, H., 2018. Pupil dilation as an index of effort in cognitive control tasks: a review. Psychon. Bull. Rev. 25 (6), 2005–2015. https://doi.org/10.3758/s13423-018-1432-y.
Van Gerven, P.W.M., Paas, F., Van Merriënboer, J.J.G., Schmidt, H.G., 2004. Memory load and the cognitive pupillary response in aging. Psychophysiology 41 (2), 167–174. https://doi.org/10.1111/j.1469-8986.2003.00148.x.
van Gog, T., Scheiter, K., 2010. Eye tracking as a tool to study and enhance multimedia learning. Learn. Instr. 20 (2), 95–99. https://doi.org/10.1016/j.learninstruc.2009.02.009.