# Investigating the Potential of EEG for Implicit Detection of Unknown Words for Foreign Language Learning

Christina Schneegass, Thomas Kosch, Albrecht Schmidt, and Heinrich Hussmann

LMU Munich, 80337 Munich, Germany
{christina.schneegass, thomas.kosch, albrecht.schmidt, heinrich.hussmann}@ifi.lmu.de

**Abstract.** Ubiquitous technologies change the way we learn new languages. They provide easy access to multilingual media content within everyday scenarios to enhance language skills or improve vocabulary. The detection of learners' vocabulary deficiencies is an integral part of effective language learning. However, retrieving translations during everyday media consumption causes inattention and hinders fluent learning. In this paper, we investigate Electroencephalography (EEG) to assess single-word incomprehension within non-native learning contents. In a user study (N=10), we employed Rapid Serial Visual Presentation (RSVP) to display text while recording Event-Related Potentials (ERPs) for different word difficulties. Our results show that incomprehension of words can be detected in participants' neural responses, as is confirmed by post hoc ratings. We conclude with use case scenarios in which our approach can facilitate seamless vocabulary deficiency detection in everyday life.

**Keywords:** EEG · Language Proficiency Detection · Implicit Interaction.

## 1 Introduction

We often encounter foreign media contents in our daily lives: when we read international news articles on social media, when we do not want to wait for dubbed versions of movies, or when we are on holiday and business trips in foreign countries. Although we can rely on today's technology to provide us with instant translations [17], learning and speaking another language is still important for our communication skills [20, 48].

However, language learning is time-consuming and highly individual. To provide efficient learning support, we need to adapt the content to target each person's specific knowledge gaps [54]. One way to discover these gaps is by continuously monitoring the users' understanding when engaging with second language content [5]. When a problem occurs, we can ensure thorough comprehension
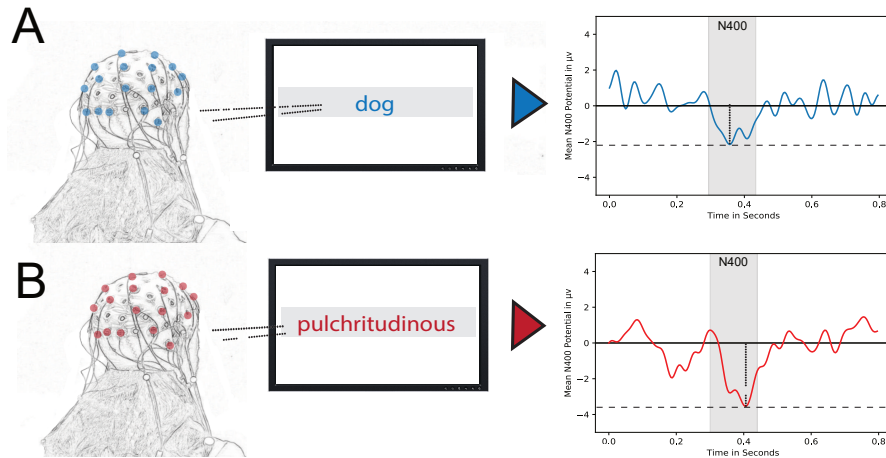
**Fig. 1.** Our concept entails the recording of EEG data while the user is reading texts including known words as opposed to unknown words on a screen. Encountering (A) known words results in a weaker N400 amplitude compared to (B) unknown words.

by interventions with either explicit or implicit user input. Explicit input, for example looking up a word in the dictionary, requires time and causes distraction. By using implicit data to detect problems in understanding, we can create interfaces, which support the user without interrupting the learning process.

In this paper, we investigate the potential of Electroencephalography (EEG) for the implicit detection of gaps in users' vocabulary knowledge for learning foreign language contents. In particular, we evaluate Event-Related Potentials (ERPs) [7] to differentiate between known and unknown words in English second language reading (see Figure 1) along with a native language baseline. In our experiment, we presented the texts as a Rapid Serial Visual Presentation (RSVP) approach, which displays text as one word at a time [46].

In a within-subject user study (N=10), we specifically looked into N400 ERPs, which are known to be indicators for syntactic processing, during the reading of native and second language text [50]. We included words, which are known and unknown to the participants and verified the participants' understanding via post hoc ratings. Based on these results, we conclude our work with the presentation of three use case perspectives for EEG based assessment of foreign language proficiency and learning support.

By being able to continuously assess language proficiency during media consumption, we are able to build language-aware interfaces to customize and optimize language learning support. Keeping in mind the rapid development of consumer-grade EEG devices, such interfaces will soon be able to provide a method for constant and unobtrusive assessment of language skills in everyday scenarios. Although we focused on the application of text reading in language

learning, this method is not limited to this domain and might be suitable in other contexts.

The contributions of this paper are threefold:

1. We show the feasibility of EEG as a method to continuously monitor comprehension during second-language text reading.
2. Based on our results, we discuss recommendations on how to facilitate this implicit input mechanism to support language learning in everyday scenarios.
3. We present potential future personalized and optimized language learning applications.

## 2  Related Work

Our work is based on previous research in the following two areas: (1) applications for language learning with media in everyday environments and (2) the cognitive processes behind EEG readings as well as their evaluation in terms of language processing and learning.

### 2.1  Language Learning in Everyday Scenarios

Learning through media and encountering foreign languages on a daily basis can have advantages that exceed those of classical classroom learning. This is mainly due to the presentation of contents in context [62]. But having context is not necessarily sufficient for comprehension. Thanks to ubiquitous computing devices, we can engage with contents and more importantly, request translations anytime and anywhere [17].

However, interrupting everyday tasks (e.g., learning vocabulary by watching a Spanish TV show) to receive a translation for a specific word via our smartphone, can be distracting from the content itself. This phenomenon, named media multitasking, can have a negative effect on recall and comprehension due to higher cognitive load in this situation [65].

Besides active user interventions, certain applications gather implicit user data to assess language comprehension and support learning. For example, language proficiency is reflected in users' gaze features. Karolus et al. (2017) [35] evaluated language proficiency for texts shown on public displays by analyzing fixations and blink duration of users. This novel approach shows the potential of eye tracking as assessment for language proficiency and experienced workload of users [40, 41]. For the assessment of advanced learners, Berzak, Katz & Levy (2018) [5] go one step further and prove the feasibility of gaze information to assess the degree of language comprehension. In a setup where English-second-language learners were presented with one-liners, the system could accurately predict participants' comprehension scores when compared to standardized tests [5]. However, the eyes respond with similar movements to various linguistic events and do not provide insights into more specific language processes of the brain [53].

Another promising method to implicitly estimate users' understanding during learning is Electroencephalography. The analysis of EEG data can reflect users' engagement, workload, attention, vigilance, fatigue, error recognition, emotions, flow, and immersion [23]. The following section will give a detailed description of this technique and its facets that are relevant to language processing.

## 2.2   Electroencephalography

Electroencephalography (EEG) is a noninvasive technique to measure electric potential from the brain by placing conductive electrodes on a person's scalp [3]. In contrast to other imaging techniques such as functional magnetic resonance imaging (fMRI), EEG hardware is comparably cheap and accessible for non-medical researchers and has a high temporal resolution [69]. EEG can detect brain responses within milliseconds of the stimulus presentation, thus, making it a feasible tool to monitor implicit reactions to learning contents [3]. The electrodes can measure potentials from $1\mu v$ up to $100\mu v$ (microvolts) in relation to a reference point, which is an additional electrode attached to the scalp or earlobe [21]. The two main features of EEG data that are frequently evaluated are (1) changes in frequency bands and (2) Event-Related Potentials (ERPs). ERPs are waveforms, representing positive and negative voltage fluctuations of the brain (an example ERP is shown in Figure 1). ERPs are responses to a given stimulus [70] originating in either sensory, motor, or cognitive events [43] (e.g., moving a hand or hearing a sound). The evaluation of ERPs can, besides others, infer the amount of processing resources [37] or cognitive load [58] in a given moment. Due to the high temporal resolution of EEG in general, this measurement allows for the differentiation of responses to individual words or speech sounds. Therefore, EEG can be applied as a method to examine differences in real-time first and second language processing [12]. Unfortunately, movements of the eyes and head, as well as muscle contractions, can cause noise in EEG data. Therefore, these measurements require additional effort to reduce extraneous influences and eliminate noise artifacts [11].

Research has shown that EEG can be applied for various purposes, exceeding its original application in medical research [19]. Due to advancements in both data classification software and EEG hardware quality [51], current technologies are becoming an increasingly robust tool for brain response evaluations even in real-world scenarios [39]. Within the last decade, research achieved important improvements in building small, wireless, and low-cost EEG devices [15]. Debener et al. (2015) [14] even built a portable EEG with printed electrodes using a smartphone for signal delivery and acquisition, and proved its feasibility for comfortable EEG signal capturing over many hours [14]. To support natural communication, Bleichner et al. (2015) [6] explored a nearly invisible EEG setup through the integration of electrodes into a baseball cap while reliably recording P300 ERPs [6]. These advancements show the potential of EEG to be applied in everyday scenarios.

### 2.3   Event-Related Potentials

ERPs can be used to evaluate brain activity during second language processing. These potentials are averaged responses from a group of trials as a reaction to a given experimental stimulus. The assumption is that a certain electrical potential occurs at a consistent time after the presentation of a stimulus to the participant [12, 25] (e.g., a non-word or a word out of context).

Due to the high temporal resolution, ERPs can reflect responses occurring within a few hundred milliseconds after the stimulus is presented. Thus, ERPs can provide insights about the processing of individual words within sentences. Since the potentials of consecutive words can overlap, serial presentation of words can foster effective detection of ERPs. Either slow rate serial presentation or artificial separation of the words can maintain the correct mapping of stimulus and response [12]. Experiments showed that semantic relationships within sentences are shown in N400 potentials: negative-going brain potential between 250-500ms after the presentation of any potentially meaningful stimulus, peaking at around 400ms [45]. However, the exact onset of the response is depending on the reading speed and can occur earlier if the speed is close to normal [42]. N400s can reveal the strength of the semantic relationship of words in context, as Kutas & Federmeier (2000) [44] highlight in their review, and the difficulty of semantic integration [50]. In addition, N400s show that there is an influence on semantic processing mechanisms when comparing first and second language processing [2, 26]. A study by Holcomb & Neville (1990) [32] found that N400s increase immediately after the processing of non-words. N400s also represent the plausibility of verb-object combinations and appear with larger amplitudes when a word is unknown, used inappropriately, or is a pseudoword [4, 45]. In addition, the amplitudes of N400 potentials reflect how expected a word is within a sentence [44, 45]. Since related work performed studies on the processing of non words [32], unexpected words [44], or words out of context [45] sentences are often incoherent, or words sound unnatural. We adapted the study setup so that it would approximate our use cases by using coherent texts from a validated source and included real English words unknown to the participants. In our case, *unknown* words are those the participants were not able to translate to their native language.

In summary, related work showed the importance and potential of implicit measures to assess language proficiency for real-world applications. The use of EEG data is gaining popularity due to reduced acquisition costs and technical research proving the applicability in everyday settings. In particular, the analysis of N400 ERPs shows potential to assess the language proficiency of individuals.

## 3   Methodology

Based on related work, the main goal of this work is to answer the following research question:

**RQ:** Can we detect gaps in users' second language vocabulary knowledge during text reading by analyzing the amplitudes of the resulting N400 potentials?

**Table 1.** Overview of the Lexile measures [47] for E1 and E2 in both their original and revised versions utilized for the study.

|  | $E1_{orig}$ | $E1_{rev}$ | $E2_{orig}$ | $E2_{rev}$ |
|---|---|---|---|---|
| **Total Sentence Length** | 29 | 29 | 24 | 24 |
| **Lexile Measure** | 600L - 700L | 900L - 1000L | 1000L - 1100L | 1000L - 1100L |
| **Mean Number of Words per Sentence Length** | 14.96 | 15.34 | 17.36 | 17.36 |
| **Total Word Count** | 419 | 445 | 434 | 434 |

To answer the research question, we conducted a lab study in which participants were required to read texts on a computer screen while recording their EEG signals.

### 3.1   Text Difficulty Selection

For this study, we included texts from the corpus of the *Asian and Pacific Speed Readings for ESL Learner* [56]. These texts include predefined English language texts on topics related to Asia and the Pacific with a supplementary set of ten single-choice comprehension questions per topic. The texts was specifically chosen because it features frequent words and easy grammar [57] to be easily understandable. We chose to include excerpts from three texts ("Life in the South Pacific Islands", "Buddhism", and "Hong Kong") and translated the first one (further termed *N1*) into the participants' native language to serve as a baseline. N1 included 30 sentences and in total 452 words, which we split into two texts of 15 sentences each. The presentation of either subset was randomized among participants to avoid effects caused by the content.

The second and third text, named *E1* and *E2*, were in English, the participants' second language. E1 contained 29 and E2 24 sentences (~450 words per text, for more details see Table 1) to generate a sufficient set of trials while not straining the user. The two texts E1 and E2 were randomly assigned to the participants for a within-subject design. We revised each text to contain ten sentences with one uncommon word (e.g., "adscititious"), selected with the help of a thesaurus and a list of unfamiliar words[1]. Including just one difficult word per sentence creates a realistic scenario and prevents the overlapping of ERPs. In regards to the changes performed in the texts, we adapted the comprehension

---

[1] www.en.oxforddictionaries.com/explore/weird-and-wonderful-words - last access 2019-06-03

questionnaires for E1 and E2. Each question is meant to check the understanding of one sentences containing a potentially unknown word.

To confirm the difficulty level of the texts, we used the *Lexile Analyzer*[2]. This tool analyzes texts and provides an approximate reading level for it based on the metrics (1) word commonness, which is reported to correlate highly with text difficulty, and (2) complexity of syntax [47]. The Lexile score can range between 200L (L for Lexile) for beginner reading, up to 1700L for advanced texts [47]. Table 1 specifies the Lexile Measures for E1 and E2 in the original version and a revised version that includes unknown words. It can be seen that the effect of difficult words on the Lexile score is only noticeable in E1, since E2 already includes many proper names and consists of longer sentences. Since the Lexile Analyzer only supports English texts, we further confirm the understandability of our texts through subjective post-hoc ratings. Participants had to answer comprehension questions as well as specify every word which they could not translate.

### 3.2   Text Presentation

We presented the texts in a Rapid Serial Visual Presentation (RSVP) mode, showing each one word at a time on the screen to reduce saccadic eye movements during normal reading behavior [59] (see Figure 2). A decrease of eye movements will lead to reduced noise in the data generated by the muscles around the eye. Furthermore, since RSVP only displays one word, the matching of the EEG signal to the dedicated stimulus can be easily performed. The word presentation rate was set to 170 words per minute (WPM) based on the findings of [8], who showed that participants' reading comprehension is best at speeds ranging from 171 to 350WPM. We decided to set the speed to the lower end of this spectrum since our participants are non-native speakers and to minimize overlaps in the signals due to the processing of consecutive words (cf. Section 2.3).



**Fig. 2.** In the RSVP approach, text is displayed each one word at a time. The words are centered to minimize eye movements.

---

### 3.3   Apparatus

Our setup consisted of a display (Dell U2715H; 27 inches; 60 Hertz refresh rate) and a Brainvision Liveamp[3] EEG device comprising a sampling rate of 500 Hz. The EEG device provides a bandpass filter ranging from 0.1 to 1000 Hz and does not include a notch filter. Electrodes were placed in accordance to the International 10-20 layout (ground electrode: $Fpz$, reference electrode: $FCz$; see Figure 3).

### 3.4   Procedure

At the beginning of the study, we welcomed the participants and handed them a detailed study description. The description contained the study motivation and goal as well as its process. Every participant signed a consent form and randomly picked and crossed out an ID from a prepared sheet of possible IDs to assure adequate anonymity. Then they filled in a short demographic questionnaire asking for age, gender, highest educational degree, vision impairment, and neuronal diseases or disorders. Furthermore, they were asked to set all electronic devices into flight mode so as not to influence the data recording.

We carefully explained the EEG system, measured participants' head circumference to select between four different actiCAP sizes (54cm, 56cm, 58cm and 60cm), and instructed them to put on the cap. Afterward, we attached 32 electrodes (plus 2 reference electrodes) to the participants' scalp using the actiCAPs' designated 10/20 positioning system [33] (the electrode layout can furthermore be seen in Figure 3). We increased the conductivity of all 34 electrodes with high viscosity electrolyte gel and examined their impedance for reliable performance. The experiment started when the impedance of all electrodes reached the threshold of 10 $k\Omega$. The signals were recorded in a quiet, dimly lit experimental room equipped with a desk and a comfortable chair, around 80cm away from the screen.

At first, participants had to read one subset of the N1 baseline text (15 sentences). Successively, participants had to read one of the two English texts E1 and E2. In total, we recorded participants reading 29 (E1) / 24 (E2) sentences, generating recordings of, depending on the randomization, between 434 and 445 individual words per participants (see Table 1).

To evaluate how much additional perceived workload was induced by the second language texts, we presented each participant with two NASA Task Load Index (NASA-TLX) questionnaires [28, 27], one after the baseline and one after the foreign text. Both texts were followed by a comprehension test consisting of ten questions designed to target the understanding of the sentences that included unknown words. Furthermore, to confirm that participants could not translate the words that were meant to be unknown, we presented them with a print-out version of the text. We asked them to highlight all the words, which they cannot translate to their native language. In summary, for both texts the following procedure was applied:

---

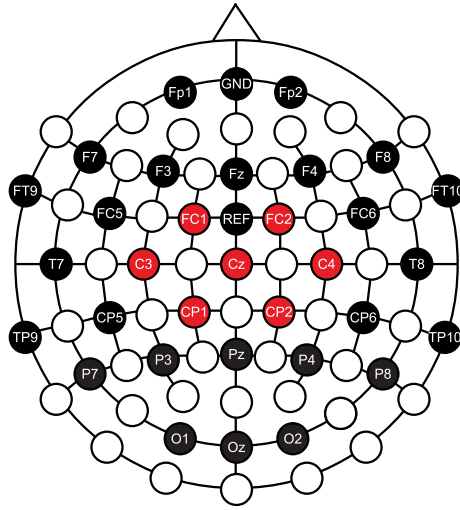[3] www.brainproducts.com - last access 2019-06-03

**Fig. 3.** The red electrodes located around the parietal lobe were used for analysis.

1. Read text as RSVP
2. Answer NASA-TLX for this text
3. Fill in ten item comprehension questionnaire
4. Post hoc rating of unknown words in printed text
5. Short rest phase

### 3.5   Sample

We recruited twelve participants via a university mailing list and an internal social media channel. As a requirement, we asked for our German and English proficiency. A minimum English proficiency of B1 according to the Common European Framework of Reference for Languages (CEFR)[4] was given due to the standards of the german high school diploma. Furthermore, we did not include participants with severe vision problems and neurological disorders. Every study participant was rewarded with an Amazon voucher. We removed two participants from our evaluation due to technical difficulties.

Within our adjusted sample size of N=10 (5 female, 5 male), the participants' age ranged from 18 - 60 ($M = 31.6$, $SD = 14.41$). They held at least a high school degree (6), some even a master degree (3), or a doctoral degree (1). Due to high school being the lowest minimal educational level of our sample, we can assume a minimum English proficiency level of B2 [18] or more for every participant.

---

[4] www.coe.int/en/web/common-european-framework-reference-languages/level-descriptions - last access 2019-06-03

### 3.6   Data Processing

We use Python with the library MNE to process the recorded EEG data[5]. EEG data were bandpass filtered [55] (0.5-40 Hz) to attenuate the influence of artifacts (e.g., blinks, eye, and head movement) as well as the 50 Hz remote power line noise. We consider the electrodes *Cz*, *C3*, *C4*, *CP1*, *CP2*, *FC1*, and *FC2* as the parietal lobe is linked to the processing of spoken and written language [68]. We identified eye blinks using Python MNE and removed them manually. We did not perform an independent component analysis as we employed 32 electrodes and we did not intend to remove the contribution of cortical components that might have been resolved into the summed activity of non-cortical dipoles. To analyze ERPs independently from known and unknown words, we slice the data set into triggers for known and unknown words. We look at the first second of neural responses for each word as we are interested in investigating the N400 ERPs that occur between 300 ms and 600 ms after displaying the stimuli.

## 4   Results

We statistically analyzed the collected data for differences in ERP magnitudes. We submitted the magnitudes of the averaged N400s for known and unknown words to an analysis of variances (ANOVA). Furthermore, we investigated the subjectively perceived workload and reading comprehension. Our results contain the EEG responses to 4390 words, of which 100 are classified as likely to be unknown to the users.

### 4.1   Event-Related Potentials

We divided the measured data into each epoch for known and unknown words. Each epoch has the same duration as a single word is displayed on the screen. We averaged each epoch for every participant and normalized the magnitude of the data to enable person-independent comparisons for native, known, and unknown words. Mauchly's test did not show a violation of sphericity. A repeated measures ANOVA was performed including the three conditions native, known, and unknown words as independent variables and the N400 amplitudes as depending variable. The analysis revealed a significant main effect ($F(2, 18) = 13.33$, $p < .001$). A post hoc test using a Bonferroni correction revealed a significant effect in N400 potentials between known and unknown words ($p < .001$, $d = 2.648$) as well as unknown and native words ($p = .031$, $d = -1.024$). We found no significant effect between the amplitudes of native and known English words. Figure 4 shows the averaged N400 across all participants for known and unknown words. The mean amplitude for known words was higher ($M = -1.201$, $SD = 0.662$) compared to the mean amplitude of unknown words ($M = -2.885$, $SD = 0.057$). Figure 5 illustrates the difference of the N400 magnitudes for known and unknown words.

---

[5] www.martinos.org/mne/stable/index.html - last access 2019-06-03
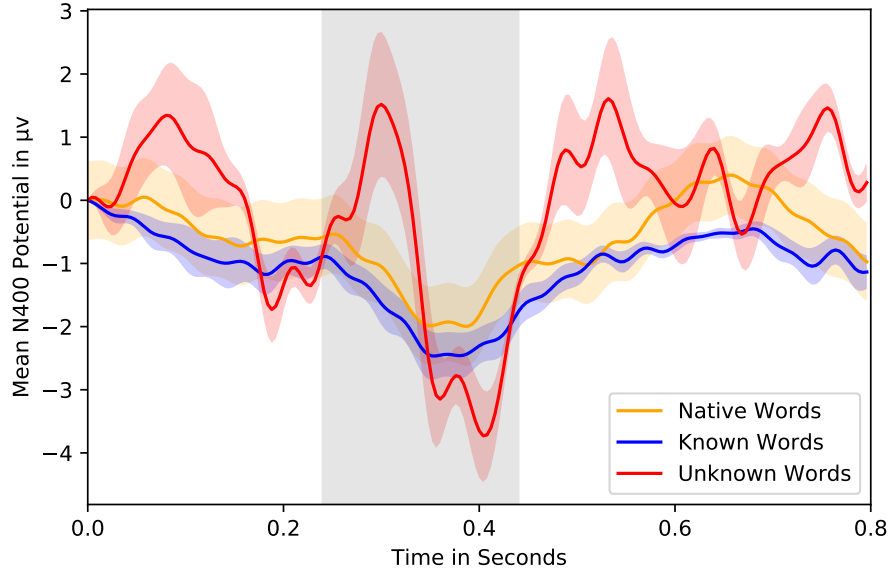
**Fig. 4.** N400 measured for known and unknown words. A larger mean amplitude is measured for unknown words compared to known words.

## 4.2  Perceived Workload

The workload during the reading of native text was perceived as lower as during English texts. The NASA-TLX is subdivided into six facets of workload: mental, physical, temporal, performance, effort, and frustration [28]. The NASA-TLX was presented as a detailed scale ranging from one (very low demand) to 20 (very high demand). For a general analysis, we added up the individual facets of the NASA-TLX to create one overall score, which therefore had a range from 1 to 120 ($6 \times 20$, the max value of one facet). This score was lower for native language text ($M = 40.4$, $SD = 19.08$) as opposed to the foreign text ($M = 53$, $SD = 18.66$). We performed a paired samples t-test comparing the overall perceived workload of the two languages (native vs. English). The results revealed a significant difference between the languages ($p < .05$, $t(9) = 2.842$, $d = -0.899$), showing a large effect in the direction of a lower mean workload in the native language texts. A Shapiro-Wilk test showed no indication for a deviation of normality ($p = .915$).

Moreover, paired samples t-tests within the individual facets showed significant differences between the two languages in terms of perceived mental workload ($t(9) = -3.452$, $p < .05$, $d = -0.506$), perceived temporal demand (i.e., feeling rushed; $t(9) = -2.339$, $p < .05$, $d = -0.740$), and perceived performance (i.e., reading and understanding the texts; $t(9) = -2.872$, $p < .05$, $d = -0.899$). All of these results showed negative values for Cohen's d, indicating higher loads for

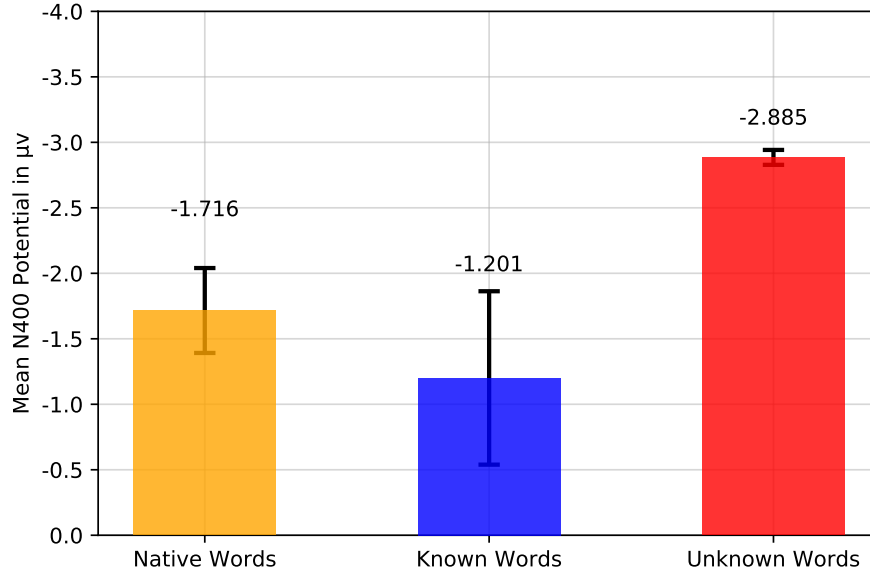the English texts. Figure 6 shows the mean raw NASA-TLX score between both languages.



**Fig. 5.** Mean N400 amplitudes for known and unknown words. Unknown words elicit a statistical significant effect in amplitudes compared to known words. The bars depict the standard error.

### 4.3   Text Comprehension

There was each one questionnaire to test the comprehension of the N1, E1, and E2 texts. On average, the participants achieved the best scores in the N1 questionnaire with around 7 correct answers (correct answers out of ten, $M = 6.9$, $SD = 2.13$), followed by E2 ($M = 6.6$, $SD = 1.82$). The text with the least amount of correct answers was E1 ($M = 5.4$, $SD = 2.19$).

### 4.4   Post Hoc Word Review

Participants performed a post hoc rating of all words, highlighting each word for which they can not come up with a translation. No participant highlighted any word of the N1 text. We can furthermore confirm the low difficulty of the E1 and E2 due to the fact that not a single word of all three texts was perceived as difficult besides our artificial modifications. When looking at the 10 modified words of E1 in detail, the results show a high consensus among the participants

who read the text. Out of the 10 potentially unknown words, 7 are confirmed as unknown by all five participants ($M = 4.6$, $SD = 0.7$). In the text E2, the confirmation of the unknown words turned out to be less distinct. Only three out of the 10 words were highlighted by all participants. On average, the potentially unknown words are highlighted by 4.2 participants ($SD = 0.63$).
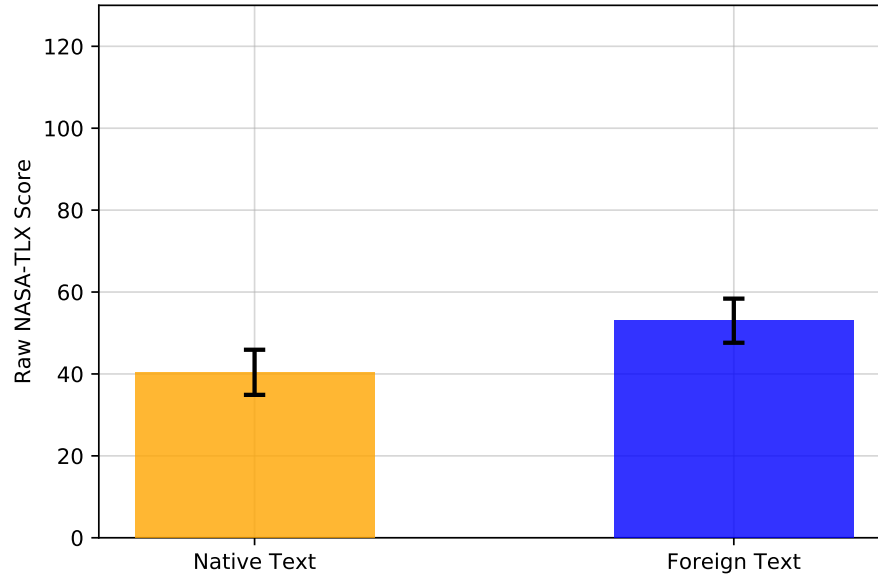


**Fig. 6.** Mean raw NASA-TLX scores for both languages. Reading native languages resulted in less workload compared to foreign languages. The bars depict the standard error.

## 5  Discussion

We conducted a user study to investigate the feasibility of ERPs to detect vocabulary gaps. Our results show a statistically significant main effect in N400 amplitudes between known and unknown words as well as between native and unknown words. In the following, we discuss the implications of our results.

### 5.1  Detecting Vocabulary Gaps

The results from our study show that we can use EEG data, N400 ERPs in particular, to assess the word-based language proficiency by measuring a significant effect between the amplitudes caused by known and unknown words. This confirms EEG as a valid tool for vocabulary gap detection. Although the descriptive

ERP data showed minor differences between the N400 amplitudes of reading native words and reading known second-language words, the statistical analysis did not result in a significant difference. We conclude that N400s, independent from the presented language, have common properties [66]. Going beyond the results of the conducted study, our results encourage further investigations on neural activity of second language processing.

## 5.2  Subjective Workload and Comprehension

When comparing the results of the raw NASA-TLX questionnaires we recognize a significant difference in perceived workload when reading native as opposed to English words. This shows that subjectively perceived workload was manipulated in conjunction with our finding in N400 amplitude. The comprehension tests show that there is a difference of complexity when comparing the two English texts. However, when marking the unknown words in the print-outs, the participants reached a 70% consensus. Therefore, we assume that the participants perceived the unknown words as equally difficult to translate in both English texts. We infer that the text comprehension rate is affected by the text difficulty, however, this does not have an impact on the individual N400 measures.

Participants achieved lower text comprehension scores for the text E1 compared to E2 although E2 is more difficult according to the Lexile score. Since the Lexile scores takes several syntactic and semantic factors into account to calculate the text difficulty, a potential threat to validity is posed by participants that were familiar with the text. However, we have observed the effect of N400 for most of the words that were unknown for participants. Thus, we believe that the overall text difficulty does not represent a higher occurrence of N400s.

## 5.3  Differences in other ERP characteristics

Besides the characteristic N400 amplitude differences, Figure 4 revealed a set of further differences in the EEG signals. Reading foreign known and unknown words seems to reflect in increased positive amplitudes at around 100ms as well as around 300ms after stimulus onset. In addition, the signal received for unknown words shows a higher mean potential at around 500-600ms after the stimulus. The latter could reveal a P600 [43] induced by syntactic continuation problems or checking upon unexpected (linguistic) events [38]. The P600 is related to the P300 [10], which can be a result of the 'oddball' effect. The oddball effect is a phenomenon of inattention blindness and can occur if an unexpected stimulus appears [64]. In our case, the unknown words suddenly interrupted the fluent reading behaviour. It is common, that a P300 occurs simultaneously with an N200 [10]. Further statistical analysis need to evaluate the differences of other ERP components in the recorded signals.

### 5.4   Limitations

A major challenge of this approach is the sum of influencing factors, which would reflect in the EEG data when applied in a real-world setting. We minimized these effects by conducting a study in a laboratory setting and were able to control people's attention and task load. Therefore, we do not know to what extent our results are generalizable to other situations. Including additional measurements such as gaze tracking could help to compensate for other influences in the application of EEG data in everyday settings. Furthermore, we have to examine in small steps the potential of this approach when faced with further stimuli (e.g., by adding video or auditory material). The applicability for other text presentation modes, for example including a sentence-based text presentation mode as an approximation of subtitles used in videos, needs to be evaluated. Furthermore, we acknowledge that our study employed a low sample size. However, we replicated the methodology from other HCI studies that have successfully employed similar sample sizes [29, 63] and therefore, believe that this study can highlight the potential of EEG for implicit language proficiency detection. We see our work as a first proof of concept and as to be following the path of other EEG research publishing novel ideas for real-world scenarios [30, 60].

## 6   Use Cases for EEG as Ubiquitous Language Learning Support

In the following, we present potential use case perspectives to show the application of EEG data on vocabulary comprehension to support language learning in everyday environments.

### *Use Case 1 - RSVP Reading on Small Screen Devices*
Designing efficient reading interfaces on devices with limited screen space, such as a smartwatch, is challenging within mobile contexts [16]. The usage of RSVP, where the whole screen presents a single word, is one interface design option that has been evaluated [24]. It has the advantage of presenting text in a reasonably large font and allows reading with little eye movement. However, this interface has one inherent limitation: Being presented with just one word at the time, a user cannot take a step back when encountering problems of understanding due to unknown words. With the insights from analyzing the EEG data, we would be able to detect in real-time when the user encounters unknown words or potentially troublesome concepts during reading as illustrated in Figure 7a. By combining a smartwatch and a mobile EEG device we could use an algorithm to dynamically adapt the content and interface. For example, offer real-time support by showing words for a longer time, provide translations, or by show unknown words more frequently in other media contents. The realization of this scenario is, however, depending on the development of portable and pervasive EEG devices, which have the potential to become affordable [15], unobtrusive [14], nearly invisible [6], and feasible for applications including natural actions and cognition [13].
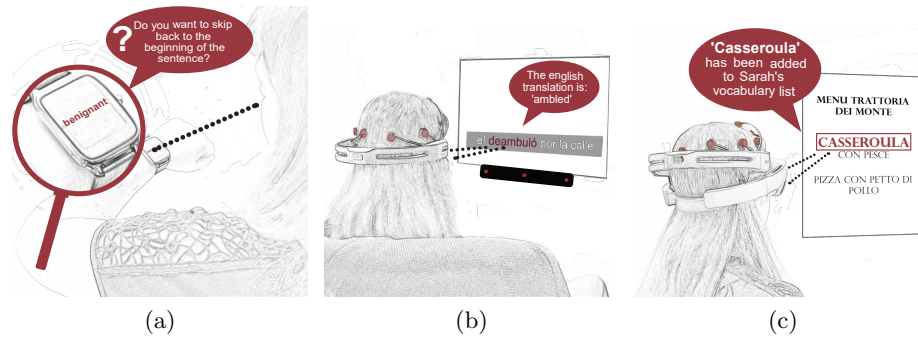
**Fig. 7.** Possible use case scenarios for EEG based language comprehension **(a):** Reading on small screen devices. **(b):** Media consumption on a screen with subtitle reading enhanced with gaze-tracking. **(c):** Reading media content in a real-world environment with supplementary smart glasses.

### Use Case 2 - Media Consumption on Screen

Using media content in foreign languages has been often reported to be a useful tool to improve language skills [31]. One area, where our approach can be beneficial, is the presentation of subtitles in videos. Including subtitles in audiovisual content can support the acquisition and the improvement of language skills [67, 52]. There are already tools exploring the potential of subtitle translations (GliFlix [61]), or second screen application to present important concepts of TV shows (*Flickstuff* [36]). By using our EEG based approach in monitoring users' comprehension during media usage as shown in Figure 7b, we can provide an effective tool for real-time and post hoc vocabulary learning support, such as personalized vocabulary lists. To implement this, one would couple the EEG monitoring and analysis with gaze tracking to detect the current focus of the users and thus, to identify the word in question.

### Use Case 3 - Media Content in Ubiquitous Environments with Smart Glasses

What we envision for subtitles on a screen could also be transferred into real-world environments. In our everyday life, we encounter signs and texts as they are ubiquitous in our surroundings. With a setup consisting of smart glasses that include a camera, we can link gaze tracking to a mobile EEG. Thus, it will become feasible to also detect signs and texts not understood by the user on many digital and analogue devices such as advertisements or public screens. The general approach to apply comprehension analysis in the physical surrounding is to detect what the user is perceiving and assess their brain's reaction. Figure 7c illustrates our vision that this can be realized by smart glasses and a front-facing camera. With the help of optical character recognition [49], it is feasible to monitor text in the users' surroundings [22] and provide individual support.

## 7   Conclusion and Future Work

The findings from our work provide evidence for the potential of Electroencephalography (EEG) data to support language learning. In particular, Event-related Potentials (ERPs) are a feasible measurement for the detection of incomprehension on a one-word basis. Having an approach to support the continuous assessment of second-language text comprehension brings us one step closer towards the design of ubiquitous learning support. We tested the approach in the context of text reading and aimed to recognize vocabulary incomprehension to support language learning. Still, this method is not limited to this application scenario and is of particular interest to many areas of ubiquitous technology. Three possible use case scenarios were outlined in this work, highlighting the additional value an EEG based word comprehension system would offer. Additionally to the use case of foreign language reading, future work should investigate the transfer of this approach to evaluate spoken language comprehension. Thereby, this technique could support real life communication, which is, in particular, important in conversations where the two involved parties show different levels of language proficiency. Furthermore, different foreign languages may elicit different magnitudes in the measured N400 amplitudes. Investigating the difference of N400 amplitudes within different languages is subject for future work.

Although this work supports the assumption that ERPs have a high potential for vocabulary learning support, further evaluation needs to clarify the feasibility for real-time support. In general, ERPs are noisy and, therefore, need to be averaged over many trials. However, requiring the user to take a training phase and applying deep learning can be beneficial, as priorly applied for emotion classification [34, 71] or the evaluation of motor imagery signals [1, 9]. The result could be the establishment of an individual baseline. This baseline could function as a classifier to distinguish known and unknown words based on the ERP signals in real-time settings. Moreover, the RSVP approach introduces certain degrees of freedom, e.g., text presentation speed. Providing maximum speed while still being able to measure ERPs needs to be evaluated in future work. We believe that this work provides a first step on a path towards a new generation of personal assistance system based on the usage of EEG technology.

## References

1. An, X., Kuang, D., Guo, X., Zhao, Y., He, L.: A deep learning method for classification of eeg data based on motor imagery. In: International Conference on Intelligent Computing. pp. 203–210. Springer (2014)
2. Ardal, S., Donald, M.W., Meuter, R., Muldrew, S., Luce, M.: Brain responses to semantic incongruity in bilinguals. Brain and language **39**(2), 187–205 (1990)
3. Baillet, S., Mosher, J.C., Leahy, R.M.: Electromagnetic brain mapping. IEEE Signal processing magazine **18**(6), 14–30 (2001)
4. Bentin, S., McCarthy, G., Wood, C.C.: Event-related potentials, lexical decision and semantic priming. Electroencephalography and clinical Neurophysiology **60**(4), 343–355 (1985)

5.  Berzak, Y., Katz, B., Levy, R.: Assessing language proficiency from eye movements in reading. arXiv preprint arXiv:1804.07329 (2018)
6.  Bleichner, M.G., Lundbeck, M., Selisky, M., Minow, F., Jäger, M., Emkes, R., Debener, S., De Vos, M.: Exploring miniaturized eeg electrodes for brain-computer interfaces. an eeg you do not see? Physiological reports **3**(4) (2015)
7.  Borovsky, A., Kutas, M., Elman, J.: Learning to use words: Event-related potentials index single-shot contextual word learning. Cognition **116**(2), 289–296 (2010)
8.  Chen, C.H., Chien, Y.H.: Effects of rsvp display design on visual performance in accomplishing dual tasks with small screens. International Journal of Design **1**(1) (2007)
9.  Cheng, P., Autthasan, P., Pijarana, B., Chuangsuwanich, E., Wilaiprasitporn, T.: Towards asynchronous motor imagery-based brain-computer interfaces: a joint training scheme using deep learning. arXiv preprint arXiv:1808.10852 (2018)
10. Coulson, S., King, J.W., Kutas, M.: Expect the unexpected: Event-related brain response to morphosyntactic violations. Language and cognitive processes **13**(1), 21–58 (1998)
11. Cutmore, T.R., James, D.A.: Identifying and reducing noise in psychophysiological recordings. International Journal of Psychophysiology **32**(2), 129–150 (1999)
12. Davidson, D.: Brain activity during second language processing (erp). The Encyclopedia of Applied Linguistics (2012)
13. De, M.V., Debener, S.: Mobile eeg: towards brain activity monitoring during natural action and cognition. International Journal of Psychology **91**(1), 1–2 (2014)
14. Debener, S., Emkes, R., De Vos, M., Bleichner, M.: Unobtrusive ambulatory eeg using a smartphone and flexible printed electrodes around the ear. Scientific reports **5**, 16743 (2015)
15. Debener, S., Minow, F., Emkes, R., Gandras, K., De Vos, M.: How about taking a low-cost, small, and wireless eeg for a walk? Psychophysiology **49**(11), 1617–1621 (2012)
16. Dingler, T., Rzayev, R., Schwind, V., Henze, N.: Rsvp on the go: implicit reading support on smart watches through eye tracking. In: Proceedings of the 2016 ACM International Symposium on Wearable Computers. pp. 116–119. ACM (2016)
17. Doherty, S.: Translations— the impact of translation technologies on the process and product of translation. International Journal of Communication **10**, 23 (2016)
18. of Europe. Council for Cultural Co-operation. Education Committee. Modern Languages Division, C.: Common European Framework of Reference for Languages: learning, teaching, assessment. Cambridge University Press (2001)
19. van Erp, J., Lotte, F., Tangermann, M.: Brain-computer interfaces: beyond medical applications. Computer **45**(4), 26–34 (2012)
20. Fan, S.P., Liberman, Z., Keysar, B., Kinzler, K.D.: The exposure advantage: Early exposure to a multilingual environment promotes effective communication. Psychological Science **26**(7), 1090–1097 (2015)
21. Fisch, B.: Fisch and Spehlmann's EEG Primer: Basic Principles of Digital and Analog EEG, 3e. Amsterdam: Elsevier (1999)
22. Fragoso, V., Gauglitz, S., Zamora, S., Kleban, J., Turk, M.: Translatar: A mobile augmented reality translator. In: Applications of Computer Vision (WACV), 2011 IEEE Workshop on. pp. 497–502. IEEE (2011)
23. Frey, J., Mühl, C., Lotte, F., Hachet, M.: Review of the use of electroencephalography as an evaluation method for human-computer interaction. arXiv preprint arXiv:1311.2222 (2013)

24. Gannon, E., He, J., Gao, X., Chaparro, B.: Rsvp reading on a smart watch. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting. vol. 60, pp. 1130–1134. SAGE Publications Sage CA: Los Angeles, CA (2016)

25. Hagoort, P., Hald, L., Bastiaansen, M., Petersson, K.M.: Integration of word meaning and world knowledge in language comprehension. science **304**(5669), 438–441 (2004)

26. Hahne, A.: What's different in second-language processing? evidence from event-related brain potentials. Journal of psycholinguistic research **30**(3), 251–266 (2001)

27. Hart, S.G.: Nasa-task load index (nasa-tlx); 20 years later. In: Proceedings of the human factors and ergonomics society annual meeting. vol. 50, pp. 904–908. Sage Publications Sage CA: Los Angeles, CA (2006)

28. Hart, S.G., Staveland, L.E.: Development of nasa-tlx (task load index): Results of empirical and theoretical research. In: Advances in psychology, vol. 52, pp. 139–183. Elsevier (1988)

29. Hassib, M., Pfeiffer, M., Schneegass, S., Rohs, M., Alt, F.: Emotion actuator: Embodied emotional feedback through electroencephalography and electrical muscle stimulation. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems. pp. 6133–6146. ACM (2017)

30. Hassib, M., Schneegass, S., Eiglsperger, P., Henze, N., Schmidt, A., Alt, F.: Engagemeter: A system for implicit audience engagement sensing using electroencephalography. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems. pp. 5114–5119. ACM (2017)

31. Herron, C., Morris, M., Secules, T., Curtis, L.: A comparison study of the effects of video-based versus text-based instruction in the foreign language classroom. French Review pp. 775–795 (1995)

32. Holcomb, P.J., Neville, H.J.: Auditory and visual semantic priming in lexical decision: A comparison using event-related brain potentials. Language and cognitive processes **5**(4), 281–312 (1990)

33. Jasper, H.H.: The ten-twenty electrode system of the international federation. Electroencephalogr. Clin. Neurophysiol. **10**, 370–375 (1958)

34. Jirayucharoensak, S., Pan-Ngum, S., Israsena, P.: Eeg-based emotion recognition using deep learning network with principal component based covariate shift adaptation. The Scientific World Journal **2014** (2014)

35. Karolus, J., Wozniak, P.W., Chuang, L.L., Schmidt, A.: Robust gaze features for enabling language proficiency awareness. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems. pp. 2998–3010. ACM (2017)

36. Knittel, J., Dingler, T.: Mining subtitles for real-time content generation for second-screen applications. In: Proceedings of the ACM International Conference on Interactive Experiences for TV and Online Video. pp. 93–103. ACM (2016)

37. Kok, A.: Event-related-potential (erp) reflections of mental resources: a review and synthesis. Biological psychology **45**(1-3), 19–56 (1997)

38. Kolk, H.H., Chwilla, D.J., Van Herten, M., Oor, P.J.: Structure and limited capacity in verbal working memory: A study with event-related potentials. Brain and language **85**(1), 1–36 (2003)

39. Kosch, T., Funk, M., Schmidt, A., Chuang, L.: Identifying cognitive assistance with mobile electroencephalography: A case study with in-situ projections for manual assembly. In: Proceedings of the 10th ACM SIGCHI symposium on Engineering interactive computing systems. ACM (2018). https://doi.org/10.1145/3229093

40. Kosch, T., Hassib, M., Buschek, D., Schmidt, A.: Look into my eyes: Using pupil dilation to estimate mental workload for task complexity

adaptation. In: Proceedings of the 2018 CHI Conference Extended Abstracts on Human Factors in Computing Systems. CHI EA '18, ACM, New York, NY, USA (2018). https://doi.org/10.1145/3170427.3188643, https://doi.org/10.1145/3170427.3188643

41. Kosch, T., Hassib, M., Wozniak, P., Buschek, D., Alt, F.: Your eyes tell: Leveraging smooth pursuit for assessing cognitive workload. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. CHI '18, ACM, New York, NY, USA (2018). https://doi.org/10.1145/3173574.3174010

42. Kutas, M.: Event-related brain potentials (erps) elicited during rapid serial visual presentation of congruous and incongruous sentences. Electroencephalography and Clinical Neurophysiology **40**(Supplement), 406–411 (1987)

43. Kutas, M., Dale, A.: Electrical and magnetic readings of mental functions. Cognitive neuroscience pp. 197–242 (1997)

44. Kutas, M., Federmeier, K.D.: Electrophysiology reveals semantic memory use in language comprehension. Trends in cognitive sciences **4**(12), 463–470 (2000)

45. Kutas, M., Hillyard, S.A.: Reading senseless sentences: Brain potentials reflect semantic incongruity. Science **207**(4427), 203–205 (1980)

46. Legge, G.E., Mansfield, J.S., Chung, S.T.: Psychophysics of reading: Xx. linking letter recognition to reading speed in central and peripheral vision. Vision research **41**(6), 725–743 (2001)

47. Lennon, C., Burdick, H.: The lexile framework as an approach for reading measurement and success. electronic publication on www. lexile. com (2004)

48. Liberman, Z., Woodward, A.L., Keysar, B., Kinzler, K.D.: Exposure to multiple languages enhances communication skills in infancy. Developmental science **20**(1), e12420 (2017)

49. Mori, S., Nishida, H., Yamada, H.: Optical character recognition. John Wiley & Sons, Inc. (1999)

50. Mueller, J.L.: Electrophysiological correlates of second language processing. Second Language Research **21**(2), 152–174 (2005)

51. Mullen, T.R., Kothe, C.A., Chi, Y.M., Ojeda, A., Kerth, T., Makeig, S., Jung, T.P., Cauwenberghs, G.: Real-time neuroimaging and cognitive monitoring using wearable dry eeg. IEEE Transactions on Biomedical Engineering **62**(11), 2553–2567 (2015)

52. Neuman, S.B., Koskinen, P.: Captioned television as comprehensible input: Effects of incidental word learning from context for language minority students. Reading research quarterly pp. 95–106 (1992)

53. Osterhout, L., McLaughlin, J., Kim, A., Greenwald, R., Inoue, K.: Sentences in the brain: Event-related potentials as real-time reflections of sentence comprehension and language learning. The on-line study of sentence comprehension: Eyetracking, ERP, and beyond pp. 271–308 (2004)

54. Pearson, P.D., Hiebert, E.H., Kamil, M.L.: Vocabulary assessment: What we know and what we need to learn. Reading research quarterly **42**(2), 282–296 (2007)

55. Proakis, J.G.: Digital signal processing: principles algorithms and applications. Pearson Education India (2001)

56. Quinn, E., Nation, I.S.P.: Speed reading: A course for learners of English. Oxford University Press (1974)

57. Quinn, E., Nation, I.S.P., Millett, S.: Asian and pacific speed readings for esl learners. English Language Institute Occasional Publication **24** (2007)

58. Roy, R.N., Charbonnier, S., Campagne, A., Bonnet, S.: Efficient mental workload estimation using task-independent eeg features. Journal of neural engineering **13**(2), 026019 (2016)

59. Rubin, G.S., Turano, K.: Reading without saccadic eye movements. Vision research **32**(5), 895–902 (1992)
60. Sahami Shirazi, A., Funk, M., Pfleiderer, F., Glück, H., Schmidt, A.: Mediabrain: Annotating videos based on brain-computer interaction. Mensch & Computer 2012: interaktiv informiert–allgegenwärtig und allumfassend!? (2012)
61. Sakunkoo, N., Sakunkoo, P.: Gliflix: Using movie subtitles for language learning. In: Proceedings of the 26th Symposium on User Interface Software and Technology. ACM (2013)
62. Secules, T., Herron, C., Tomasello, M.: The effect of video context on foreign language learning. The Modern Language Journal **76**(4), 480–490 (1992)
63. Shirazi, A.S., Hassib, M., Henze, N., Schmidt, A., Kunze, K.: What's on your mind?: mental task awareness using single electrode brain computer interfaces. In: Proceedings of the 5th Augmented Human International Conference. p. 45. ACM (2014)
64. Stokes, T.A., Welk, A.K., Zielinska, O.A., Gillan, D.J.: The oddball effect and inattentional blindness: How unexpected events influence our perceptions of time. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting. vol. 61, pp. 1753–1757. SAGE Publications Sage CA: Los Angeles, CA (2017)
65. Van Cauwenberge, A., Schaap, G., Van Roy, R.: "tv no longer commands our full attention": Effects of second-screen viewing and task relevance on cognitive load and learning from news. Computers in Human Behavior **38**, 100–109 (2014)
66. Van Hell, J.G., Tokowicz, N.: Event-related brain potentials and second language learning: Syntactic processing in late l2 learners at different l2 proficiency levels. Second Language Research **26**(1), 43–74 (2010)
67. Vanderplank, R.: Resolving inherent conflicts: Autonomous language learning from popular broadcast television. Barriers and bridges: Media technology in language learning pp. 119–133 (1994)
68. Wise, R.J.S., Brownsett, S.L.E.: The Contribution of the Parietal Lobes to Speaking and Writing. Cerebral Cortex **20**(3), 517–523 (06 2009)
69. Wolpaw, J.R., Birbaumer, N., McFarland, D.J., Pfurtscheller, G., Vaughan, T.M.: Brain–computer interfaces for communication and control. Clinical neurophysiology **113**(6), 767–791 (2002)
70. Zani, A., Proverbio, A.M.: Cognitive electrophysiology of mind and brain. In: The cognitive electrophysiology of mind and brain, pp. 3–12. Elsevier (2003)
71. Zheng, W.L., Zhu, J.Y., Peng, Y., Lu, B.L.: Eeg-based emotion classification using deep belief networks. In: Multimedia and Expo (ICME), 2014 IEEE International Conference on. pp. 1–6. IEEE (2014)