Tracking individual auditory attention during learning via EEG-based neural envelope tracking

Davina Van den Broek[1], Elien Bellon[1], Tianyi Li[2], Simon Geirnaert[2], Alexander Bertrand[2] and Bert De Smedt[1]

1 Faculty of Psychology and Educational Sciences, Parenting and Special Education Research Group, Leuven, Belgium

2 Department of Electrical Engineering (ESAT), STADIUS Center for Dynamical Systems, Signal Processing and Data Analytics, KU Leuven, Belgium

Corresponding author: Davina Van den Broek – davina.vandenbroek@kuleuven.be

Abstract

Attention is fundamental for classroom learning, yet measuring it during learning remains challenging. Behavioural measures are often subjective and lack the sensitivity to capture online momentary fluctuations in attention. This experiment examined the potential of EEG(Electroencephalography)-based neural envelope tracking (NET) as a measure of auditory attention to the teacher's voice. Instead of measuring overall attention, NET considers where attention is directed to by correlating the teacher's voice with the EEG-decoded stimulus. EEG data were collected while participants (*n* = 30) watched a video lecture on neuromyths that included attention manipulations (e.g., distraction, enhancement). Participants provided self-reports on attention levels during the lecture and completed a performance test on its content. Our results revealed that attention fluctuations induced by our manipulations were effectively captured by EEG-based NET measures, demonstrating the validity of NET as a valid measure of auditory attention during learning.

1. Introduction

Attention plays a central role in academic classroom learning (Olney et al., 2015; Keller et al., 2020). However, measuring attention in the classroom presents significant challenges. Various methods have been developed, ranging from behavioural to physiological measures, each with its own strengths and limitations (Booth et al., 2023). Yet, a universally valid approach to measure attention in the classroom still does not exist. The present study therefore aims to test a new method for objectively tracking auditory attention in individuals over time in an educational context based on neural envelope tracking using electroencephalography (EEG) recordings (O'Sullivan et al., 2015).

Various behavioural methods have been used to measure attention during classroom learning, including self-reports, observations, and performance tests. Self-reports are typically collected through questionnaires and a common approach involves administering self-reports at the end of, for example, a lecture, allowing participants to reflect on their overall attentiveness. For instance, Hobbiss and Lavie (2024) asked participants, after a 50-minute lecture, to estimate the duration of their distraction across seven different categories (e.g., background noise, people around you, unrelated thoughts) and these estimates of distraction were predicted by performance on a visual search task that measured attention. Similarly, Blasiman et al. (2018) asked participants to rate their learning after an online lecture on a 10-point Likert scale. During the lecture, participants were instructed to engage in (distracting) activities such as folding laundry, playing a computer game, texting on a mobile. These activities resulted in participants rating their learning lower compared to the baseline condition (without such activity). In addition to the self-report, participants had to complete a performance test, which also demonstrated lower performance during distractions.

It is important to note that these self-reports at the end of a lecture require participants to assess their attention over an extended period, making it difficult to capture all fluctuations of attention during a lecture. This issue is sometimes addressed by asking participants to rate their attention at a specific moment during class, for example through probe questions or by the use of clickers. In the study by Risko et al. (2013), six probe questions were included into a 60-minute lecture. During the lecture, half of the group was asked to complete several non-lecture related computer tasks to investigate the impact of computer use during class. The lecture was paused for 30 seconds when a probe question appeared and the participants had to indicate what they were thinking about just before the probe question appeared. Participants could choose between the lecture, the time/computer (depending on the condition), or something else. In addition to the probe questions, the lectures were video recorded and attention was also rated by observers.

There was a significant overlap between attention scored through the probe questions and attention scored by video observations. Risko et al. (2013) also included a performance test at the end, and participants who indicated more mind wandering through the probe questions performed worse. However, this method only provides information about attention immediately prior to the probe question, and may disrupt periods of high attention, which is to be avoided when measuring attention (Keller et al., 2020).

A solution to this problem might be the use of clickers, and participants are asked to click whenever they notice an attention lapse (Bunce et al., 2010; Hlas et al., 2019). Bunce et al. (2010) asked students to use this method during a chemistry course to determine how long students can sustain their attention and whether this is influenced by different teaching methods. Participants received a clicker with three buttons, each representing a different duration of distraction: 1 minute or less, 2-3 minutes, or 5 minutes or more. They were asked to use this clicker and report every lapse of attention during every chemistry course over a four-week period. The study found that students' attentive periods were much shorter than 10-20 minutes, with most attention lapses lasting less than a minute. Unsworth and McMillan (2017) used a diary instead of a clicker to let students report attention lapses and their origin during lectures and study sessions. They found that students were most distracted by surrounding conversations and behaviours, and these self-reports were related to working memory, attention control and fluid intelligence measured in the lab. All these methods using self-reports rely on participants' interpretation of attention, risking inaccurate self-reporting.

Observations are another commonly used method to measure attention (Grammer et al., 2021; Keller et al., 2020; Rapport et al., 2009). In the study by Grammer et al. (2021), both observations and EEG (which will be discussed later in more detail) were employed to assess the impact of various instructional contexts (whole-group lecture, video watching, group discussion or independent work) on attention. The instructional sessions were videotaped and subsequently coded by different coders, who identified whether the student was attentive or inattentive at one-minute intervals. The observational data diverged significantly from the results based on the EEG data. For instance, group work was associated with the lowest levels of attention according to observational data, whereas the EEG data suggested higher levels of attention in this context. This shows that these two measures capture different aspects of attention. Observations require interpretation of behaviours and gaze by the observer, which is not needed for the EEG data. Therefore, observations are just as subjective as self-reports, as they depend on the observer's interpretation (Bradbury, 2016; Keller et al., 2020). Someone can appear attentive but might be thinking of something else, or vice versa.

A performance test is another behavioural method that is frequently employed as an indirect measure of attention (Bevilacqua et al., 2019; Boudewyn & Carter, 2018; Davidesco et al., 2023; Dikker et al., 2020; Zeamer & Fox Tree, 2013). In all these studies, participants are asked to complete a test with items about the content of the lecture they just attended, and higher performance scores are related to higher levels of attention. However, performance depends on many factors beyond attention, such as prior knowledge and motivation, for which reason the scores on a performance test cannot always be interpreted unambiguously.

The major limitation of these behavioural measures used to capture students' attention in the classroom is that they all struggle to capture the rapid fluctuations of attention during learning. They also tend to be subjective. Neurophysiological measures have been suggested as promising, more objective tools for assessing attention, as they allow for the monitoring of cognitive processes during learning rather than afterwards, at any moment in time, without interrupting attention by asking questions (De Smedt, 2018; Mayer, 2017). Especially electroencephalography (EEG), a non-invasive brain imaging technique with high temporal resolution, has been suggested to overcome the limitations of behavioural methods as it has the possibility to capture momentary fluctuations of attention in the classroom without being affected by subjective interpretation (Malik & Amin, 2017; Mayer, 2017).

A commonly used EEG measure of attention in the classroom is brain-to-brain synchrony, observed either between students or between students and their teacher. This phenomenon occurs due to similar activity patterns in individuals' brain during shared learning experiences, and it is enhanced with increased attention (Bevilacqua et al., 2019; Cohen et al., 2018; Davidesco, 2020; Davidesco et al., 2023; Dikker et al., 2017; Dmochowski et al., 2012; Ki et al., 2016; Poulsen et al., 2017). Dmochowski et al. (2012) investigated the potential of using brain-to-brain synchrony as a marker of attention by showing participants three videos with varying levels of arousal, each viewed twice. Their findings suggest that this brain-to-brain synchrony can indeed tell us something about the attention of people, as participants were less attentive the second time they watched the video's. Davidesco et al. (2023) examined brain-to-brain synchrony within an educational context, examining both student-student and student-teacher synchrony. In their experiment, a group of four students attended four mini-lectures, each lasting seven minutes. They found that increased brain-to-brain synchrony between student dyads in the alpha frequency band predicted better learning outcomes, assessed immediately after each lecture. Furthermore, this synchrony was predictive of learning outcomes on a delayed post-test administered one week later. Similarly, Davidesco et al. (2023) found student-teacher synchrony to be predictive for learning outcomes right after the lecture, but not with the learning outcomes on the delayed post-test. Although the EEG-synchrony measure has demonstrated its potential in measuring attention during learning, it is also related to other factors such as social dynamics and the nature of the stimuli (Bevilacqua et al., 2019; Davidesco, 2020; Dikker et al., 2017). For instance, Bevilacqua et al. (2019) found that students who reported greater social closeness to their teacher exhibited higher student-teacher brain-to-brain synchrony. Similarly, Dikker et al. (2017) observed that students who favoured a particular teaching style showed increased student-to-group brain-to-brain synchrony, which was also associated with personality traits like group affinity and empathy measured by self-reports. Furthermore, this method requires data from multiple individuals, as it is always measured between dyads or between an individual and a group. Consequently, this method cannot be used to track attention on an individual level.

To measure attention at the level of the individual with EEG, alpha power has been frequently used, with lower alpha power indicating higher attention (Boudewyn & Carter, 2018; Foxe & Snyder, 2011; Matsuo et al., 2024; O'Connell et al., 2009). Grammer et al. (2021) used alpha power to assess the impact of different instructional contexts on attention. In this study, a small group of students, two or three of whom were wearing a portable EEG device, participated in four different instructional activities: whole-group lecture, video watching, group discussion, and independent work. The results revealed that alpha power was significantly higher, indicating lower attention, in teacher-led activities compared to student-initiated activities. Dikker et al. (2020) utilized alpha power to investigate differences in attention throughout a day and over different teaching activities. They also investigated the association between alpha power and self-reports and scores on a performance test. Data were collected over several days at different times (e.g., early morning, mid-morning and afternoon). The results indicated lowest alpha power and highest performance during mid-morning classes compared to classes early in the morning or in the afternoon and during videos in comparison with lectures. Ki et al. (2016) utilized alpha power to investigate attention across different conditions. They employed both audiovisual and auditory-only stimuli, presenting each type twice: once with instructions to pay attention and once with instructions to count down from 1000 in steps of 7. The results indicated higher alpha power, indicating lower attention, in the counting condition and the auditory-only condition, as anticipated. However, when alpha power was used to predict attentional state, it did not perform better than chance, likely due to significant variability in alpha power across subjects. Another significant limitation of using alpha power as a measure of attention is that it reflects a general state of attention without indicating where their attention is directed to. For instance, a student might exhibit low alpha power during a course, suggesting high attention, but the student's attention could be directed towards something unrelated to the lecture, such as an online video.

Consequently, despite the low alpha power, the student may miss the teacher's content, leading to a misinterpretation of their attentiveness.

Research in other domains has used auditory attention decoding and more specifically neural envelope tracking (NET) to measure auditory attention to a specific stimulus (Mirkovic et al., 2015; O'Sullivan et al., 2015; Roebben et al., 2024; Straetmans et al., 2024; Vanthornhout et al., 2019). This approach might be particularly informative in studying attention in the classroom because it might be able to capture whether students are paying attention to the teacher's voice. NET relies on the assumption that brain signals synchronise with the short-term amplitude fluctuations of the speech signal (i.e., the envelope of the speech signal) when we are actively listening to a specific person, and that this coupling decreases when not attending to the speech. To quantify attention using NET, a correlation is calculated between the speech envelope and a reconstructed speech envelope from the brain responses using a pre-trained EEG decoder. In the aforementioned studies (and many others), the correlation between these two is shown to be modulated by the individual's attention to the stimulus.

This method has been employed by multiple researchers to identify which stimulus a participant is paying attention to in a competing speech paradigm (*selective* attention decoding). In these studies, participants were asked to focus on one of the speech streams while ignoring the other stimuli. The studies by O'Sullivan et al. (2015) and Mirkovic et al. (2015) asked their participants to listen to two different stories simultaneously and asked them to focus on one of the stories. Both studies revealed that the speech stream to which the participant was attending to could be reliably predicted based on NET. Straetmans et al. (2024) conducted a similar study but in more complex environments by adding background noise to the paradigm and by conducting the research in an everyday life setting, such as sitting in a hallway or walking on the street. Even in these more complex environments, it was possible to identify the attended speaker via the EEG-signal. Other studies have used NET to identify fluctuations of attention to a single stimulus (*absolute* attention decoding). The studies by Roebben et al. (2024) and Vanthornhout et al. (2019) compared conditions in which participants were asked to perform an additional task (e.g., watching a silent movie, reading a text) while listening to a stimulus, with an active listening condition in which full attention is paid to the stimulus. They found that NET was higher in the active listening condition compared to the dual-task conditions, proving that the method can also be used to identify attentional fluctuations within participants.

To the best of our knowledge, only two studies have examined NET in an educational context (Levy, Hackmon, et al., 2025; Levy, Korisky, et al., 2025). In the first study Levy, Korisky, et al. (2025), the impact of background noise on performance and NET was investigated. Participants viewed short (~40s) mini-lectures in a VR classroom while EEG was recorded. The mini-lectures where divided in three conditions: quiet, continuous background noise, and intermittent background noise. After each lecture, they answered four multiple-choice questions. Performance was slightly, but not significantly, lower in the noise condition compared to quiet condition, and significantly worse in the intermittent noise condition compared to continuous noise condition. Similarly, NET was significantly reduced in the noise conditions compared to the quiet condition, with the most pronounced reduction in the intermittent noise condition. In the second study Levy, Hackmon, et al. (2025), a similar design was used with 30 mini-lectures, 22 of which included sporadic background noise (e.g., human and artificial sounds). Participants again answered four multiple-choice questions after each lecture. The study compared attention and distractibility in individuals with and without AD(H)D. No significant performance differences were found between quiet and noise conditions, or between groups. For the NET analyses, data was only sufficient for the noise conditions, revealing slightly lower correlations in the AD(H)D group.

Evidence for the applicability of NET to assess auditory attention in educational settings is still limited. To properly evaluate its potential in classroom contexts, studies should incorporate educationally relevant stimuli to obtain results that are truly reflective of students' attentional processes during classroom learning (van Atteveldt et al., 2018). Prior research has largely relied on non-educational content, such as children stories (Mirkovic et al., 2015; O'Sullivan et al., 2015; Roebben et al., 2024; Straetmans et al., 2024; Vanthornhout et al., 2019), paradigms with competing speakers (O'Sullivan et al., 2015; Mirkovic et al., 2015; Straetmans et al., 2024), or stimuli shorter than a typical classroom lecture (Levy, Hackmon, et al., 2025; Levy, Korisky, et al., 2025). These approaches are not optimally tailored to the complex nature of education, where students are required to sustain attention and engage with educational material over extended periods.

Against this background, we examined NET to measure auditory attention to the teacher's voice in an educational context. We set up an experiment using a video lecture with a duration of 65 minutes, while EEG data were collected. During this lecture, we included multiple interventions to manipulate the attention of the participants. Specifically, we tried to enhance attention by displaying a large red exclamation mark and a red square around the video, with a prior instruction that this indicates an important part of the lecture. In order to create a distraction, a funny animal video or disturbing background noise, such as in-class background noise or construction works, was incorporated. To divert attention, we gave the participants the task to do math exercises or to read a text. We hypothesized that NET would be lower for lapses of attention caused by our distractions and dual tasks. In addition to EEG data, we also collected attention self-reports. At fixed time points during the video, we asked participants to fill in five questions about their attention during the past part of the lecture using a 10-point Likert scale. After the video, we administered a performance test in order to evaluate their knowledge about the lecture. This approach enabled the further validation of NET by examining the associations between NET and widely utilised behavioural methods. We hypothesized that our behavioural measures will be able to pick up the fluctuations in attention based on the manipulations included in the video. We further compared NET to the use of alpha power, which is a widely used metric of the overall attention in individuals. We hypothesized that NET would be more sensitive to student's attention to the teacher as compared to alpha power. Specifically, we expected lower NET during the dual-task condition, as participants would be less attentive to the teacher's voice and we did not expect alpha power to clearly reflect this drop in attention to the teacher's voice, since participants remain engaged in a task. Although a shift in alpha power is expected, its direction is unclear, as both listening and the dual task require attentional engagement.

2. Method

2.1. Participants

Data were collected from a sample of 30 participants (7 males), aged between 18 and 35 (M_{age} = 24 years; *SD* = 3 years). All participants were native Dutch speakers with normal to corrected-to-normal vision and no hearing difficulties. None of the participants reported a deficit in attention. The participants' backgrounds varied considerably in terms of their studies and careers, although nearly all were either pursuing or had completed a higher education degree.

Participants received a monetary compensation for their participation. Written consent was obtained from all participants, and the procedure was approved by the Social and Societal Ethics Committee (SMEC) of KU Leuven (G-2023-7126).

2.2. Design

The data collection consisted of two parts. First, EEG data were collected while participants watched an online lecture. During breaks in the lecture, participants answered questions about their attention (Figure 1). Second, participants were asked to complete a performance test with items about the video's content (Figure 1).

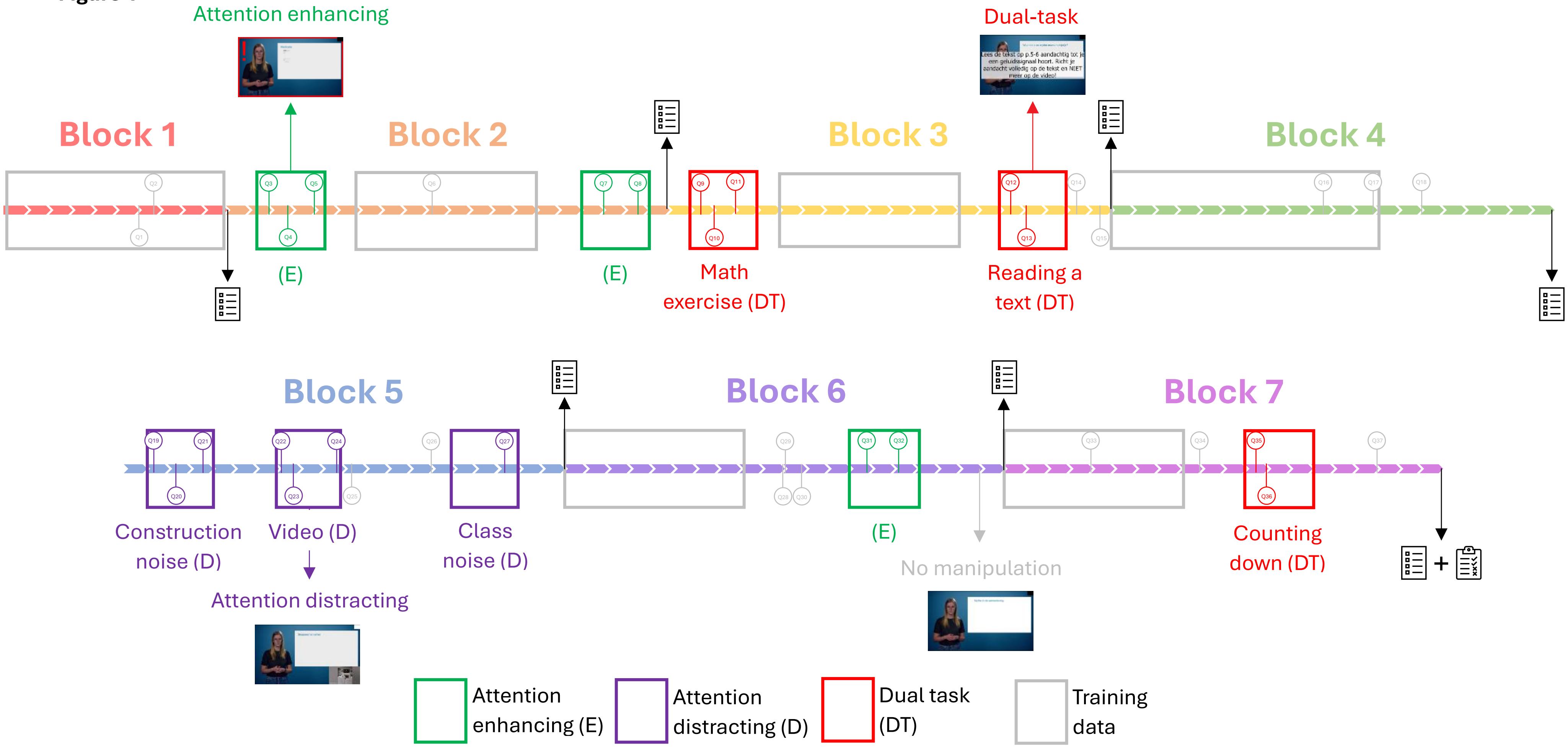
The online lecture covered the topic of neuromyths (Howard-Jones, 2010; Jarrett, 2015) and had a total duration of 65-minutes. The video depicted a teacher and a PowerPoint presentation containing only images. As the aim was to measure auditory attention, all information was given orally, with the images only supporting the spoken content. These images were not intelligible without the accompanying speech. A schematic representation of the video is provided in Figure 1. The video was divided into seven blocks: an introduction (5 minutes) and the six subsequent blocks, each addressing a different neuromyth (10 minutes each). The following neuromyths were included: (1) study pills help you study better and become smarter, (2) we only use 10% of our brains, (3) there are critical periods in child development; once these periods pass, certain things can no longer be learned, (4) teenagers who sleep a lot are just lazy, (5) differences in dominant brain hemispheres (left-brained vs. right-brained) can explain differences in performance and (6) boys have a boy's brain and girls have a girl's brain.

As shown in Figure 1 and 2, we included several manipulations throughout the video to influence participants' attention. For the attention enhancing condition, a red square appeared around the video together with a big red exclamation mark in the upper left corner. Participants were informed beforehand that they had to pay extra attention when these signs appeared, as they were sure to be asked questions from these parts. For the distracting condition, three types of distractors were used: (1) construction noise added to the audio, (2) an animal video without audio shown in the bottom right corner, and (3) classroom noise added to the audio. For the dual-task condition, three dual-tasks were given during the video: (1) completing several math exercises (e.g. 16 + 35 = ... + 47 = ... + 89 =... - 28 = ...), (2) reading a text (66 sentences) and (3) counting down in steps of 17 starting from 994 based on Ki and colleagues (2016). At the start of the experiment participants were given the dual-tasks on paper together with the instructions to do these tasks when the message appeared on the screen until they heard a beep sound. They were giving the explicit instruction to focus completely on the dual-task and to not listen to the information presented in the video during that time. Each of the three manipulation types (enhancing, distracting, dual task) lasted for 90 seconds, and each block of the video contained only one type of manipulation. The data in between these manipulations were used as a baseline condition, some of which was used to train the EEG decoding algorithm (see 2.4.1.3. Stimulus reconstruction and correlation).

Block one did not contain any manipulations and was entirely used for training the decoding algorithm (see 2.4.1.3. Stimulus reconstruction and correlation). Block two consisted of two enhancers, one in the beginning and one at the end. In between these enhancers there were four minutes that were used for training the decoding algorithm. Block three was the same as block two except with two dual-tasks instead of enhancers. Block four contained no manipulations, where the first six minutes were used for training of the decoder. Block five contained three distractors and no part that was used for training. Block six had one enhancer in the second half of the block, and the first four minutes of the block were used for decoder training. Block seven was again the same as Block six except with a dual-task instead of an enhancer. After each block participants answered questions about their attention during the past part. They received this self-report questionnaire at the start of the experiment, in the same bundle as the dual-tasks.

The EEG data were collected using a BioSemi Active Two system with 64 electrodes (10-20 system) at a sampling rate of 2048 Hz (*Biosemi,* n.d.). EEG was recorded while the participants watched the video lecture. To synchronize the EEG signal with the video, a flashing square was added to the upper right corner of the video. The square alternated between black and white every minute, which was monitored using a photosensor. This flashing square was not visible for the participants, as the photosensor was taped over the blinking corner on the screen.

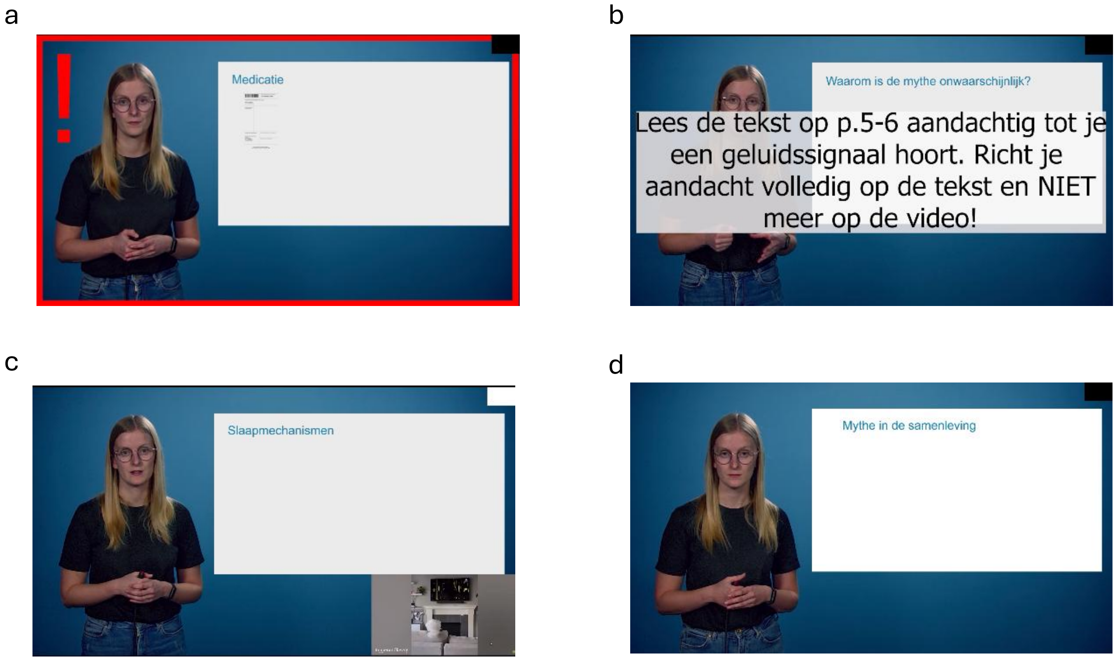
Figure 1



The figure provides a comprehensive overview of our protocol. The video consisted of 7 blocks, which are indicated with different colours, each addressing a different neuromyth. After each part the participants had to score their own attention with a questionnaire and after the video they had to do a performance test. The small (coloured) rectangles represent the nine manipulation windows and the larger (grey) rectangles indicate the parts that were used for training of the decoder. The little circles indicate which parts in the video contain the answer to a question of the performance test.

Figure 2

Overview of each condition as presented in the video



The left panel above (2a) shows the enhancing condition; the right panel above (2b) shows the dual-task condition; the left panel below (2c) shows the distracting condition and the right panel below (2d) shows baseline.

2.3. Materials

Attention self-report

A questionnaire was administered at fixed time points throughout the video to assess participants' self-reported attention levels. As previously discussed, the video was divided into seven blocks, and after each block, participants had to answer five questions regarding their attention levels. These questions were (1) How attentive were you during the previous block?, (2) How well do you think you processed the information from the previous block?, (3) How would you rate your concentration during the previous block?, (4) How long were you distracted from the lesson during the previous block? and (5) To what extent were you able to maintain your focus during the previous block? Responses were recorded on a 10-point Likert scale. To ensure that the participants read the questions carefully, the order of the questions was varied after each block.

For each block, in which one neuromyth was discussed, a mean attention score was calculated using the Likert scores on the five questions. The score for the question measuring distraction was reversed so that a higher score indicated more attention rather than more distraction.

Performance test

After they watched the video, participants completed a paper-and-pencil performance test consisting of 37 items about information presented in the video. The items were distributed equally across the different conditions with seven items pertaining to the attention enhancement condition, seven to the attention distraction condition and seven to the attention diverting condition. The remaining 16 items were from the baseline condition without any manipulations.

The test included both open and multiple-choice questions, with each condition having an equal number of open (n = 3) and multiple choice (n = 4) questions. For the baseline condition, there were seven open questions and nine multiple choice questions.

To minimize guessing, participants had the option to indicate that they did not know the answer to a question. Additionally, they could indicate if the information needed to answer the question was not present in the video. This option was included to discern whether participants did not remember the information or did not hear it. All necessary information to answer the questions was included in the video.

Participants received a score of one for each correctly answered question and a score of zero for each incorrectly answered item, as well as for those where they indicated "I don't know", or where they indicated that the information was not present in the video. The theoretical maximum score on the performance test was 37. For some analyses the percentage correct answers per condition was used as a performance measure to compensate for the difference in the number of items between the three conditions and the baseline condition.

2.4. Analyses

2.4.1. NET measure

2.4.1.1. EEG Preprocessing

In this study, we applied two stages of filtering to the EEG signal to prepare it for further analysis. First, we used a first-order IIR (infinite impulse response) high-pass filter with a cutoff frequency of 0.01 Hz to remove very low-frequency components, such as baseline drift and other slow fluctuations. After high-pass filtering, we downsampled the data to a sampling rate of 128 Hz using MATLAB (MATLAB, 2010) built-in 'resample' command (applying appropriate anti-aliasing filters).

Next, we applied a fourth-order IIR bandpass filter to focus on the key frequency components between 0.1 Hz and 4 Hz (Li et al., 2025). This frequency band was selected because it captures slow brain activities related to attention and cognition, which are crucial for phase-locking to the speech envelope, the slow variations over time of the speech stimulus. After this filtering step, the data was further downsampled to 16 Hz to improve the efficiency of subsequent signal processing algorithms.

2.4.1.2. Speech Envelope Extraction

The speech envelope served as the audio feature in this study. Prior research indicates that EEG signals can phase-lock to the speech envelope (O'Sullivan et al., 2015). To extract this feature, we employed the method from Biesmans et al. (2017), which uses a gammatone filterbank and a power law compression to simulate the frequency selectivity of the human cochlear system.

2.4.1.3. Stimulus Reconstruction and Correlation

We adopted a backward model to train a linear spatial-temporal decoder on a per-subject basis using the data in the grey boxes on Figure 1. This approach employed time-lagged versions of the EEG signals to reconstruct the speech envelope based on a least-mean-square criterion (Biesmans et al., 2017). During the validation phase, we applied the trained decoder on 30s windows and calculated the Pearson correlation coefficient between the reconstructed speech envelope and the ground-truth speech envelope. We hypothesized that a higher correlation value indicates a higher level of attention (Roebben et al., 2024). In the rest of the paper we will use the acronym NET (i.e. neural envelope tracking) to refer to the correlation between the reconstructed speech envelope and the ground-truth speech envelope, each time measured over a window of 30 seconds.

In order to further analyse the data, we average all NET values per condition obtained within the manipulation window (as indicated by the coloured boxes on Figure 1), as well as across entire blocks of the video. The latter includes a combination of data where attention was manipulated, and data where no manipulation occurred (note that each block has only one type of manipulation). This last measure, NET averaged across an entire block, is included to enable comparison with behavioural data from the self-reported attention questionnaire, in which participants rated their attention over the same time span.

2.4.2. Alpha power

To quantify alpha power, the MATLAB (MATLAB, 2010) bandpower function was used to compute the spectral power within the 8-12 Hz range (alpha band) for each of the 64 EEG channels. This analysis was performed using 30-second windows. This procedure resulted in 64 alpha power values per window. These values were than averaged to obtain a single global alpha power measurement per 30s window.

Similar as with NET, an average alpha power per condition was calculated on the one hand for every manipulation window and on the other hand for every block. These means were used in the further analyses.

2.4.3. Statistical Analysis of Behavioural and EEG Data

The behavioural data and the EEG data were further analysed using JASP (JASP Team, 2024) and R (R Core Team, 2024). To analyse the effect of our manipulations on self-reported attention, performance test scores, NET and alpha power, we used (generalized) linear mixed-effects models, given that our data was nested within participants. A generalized mixed-effects model analysis was employed for the performance test scores, given that the responses were either correct (1) or incorrect (0). For all other measures, a linear mixed-effects model was used. We examined whether the accuracy on the performance test, the self-reported attention, NET and alpha power differed between the conditions. For NET and alpha power we did this analysis when these were measured during the manipulation windows and when these were measured over an entire block.

Furthermore, we aimed to investigate the association between the behavioural measures and the EEG measures of attention. On the one hand, we looked at the correlation between the self-reported attention scores and the corresponding NET across the entire block combined per condition. Specifically, this means that for every participant we had an average self-reported attention score and NET measured over an entire block for every condition. On the other hand, we looked at the correlation between the performance test scores and the corresponding NET measured during the manipulation window for every condition. Again, we had for each participant a percentage correct answers linked with NET from the respective manipulation window. Therefore we calculated a repeated measures correlation, which is a technique to calculate the intra-individual association between two paired measures while accounting for the non-independence of the data (Bakdash & Marusich, 2017). The analyses were conducted using the rmcorr package in R (Bakdash & Marusich, 2024).

3. Results

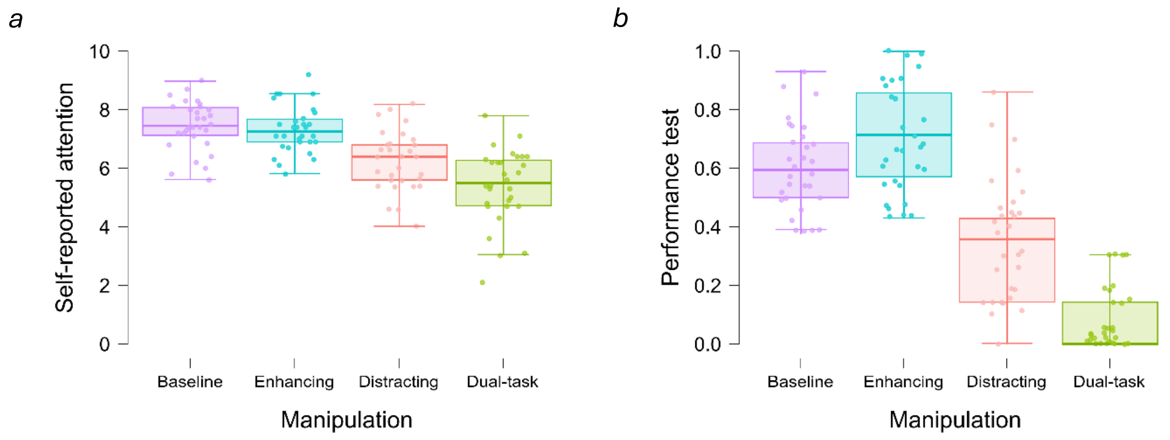
3.1. Behavioural results

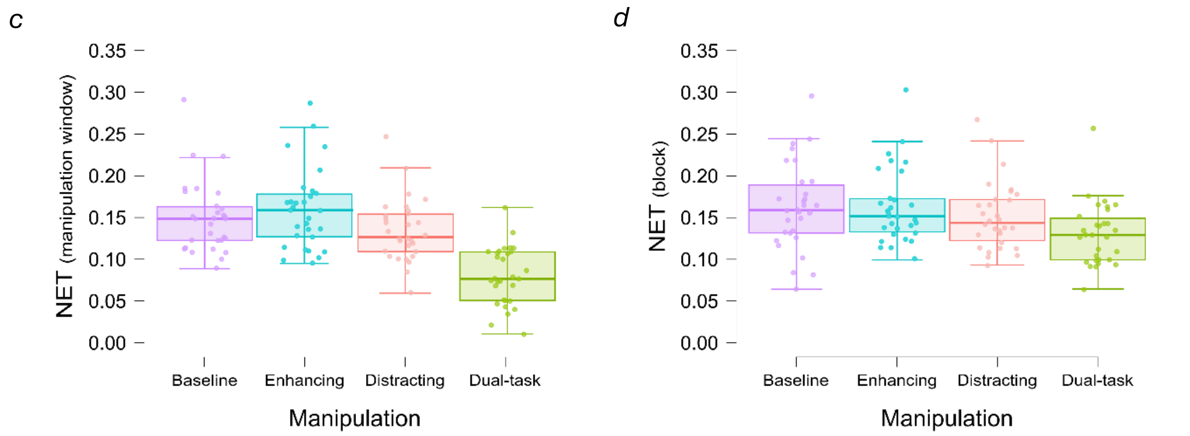
The data for the self-reported attention, separated by condition, are shown in Figure 3a. The linear mixed-effects model revealed a significant effect of manipulation on the self-reported attention scores [$F(3,177) = 51.90$, $p < .001$]. Further contrast analyses, corrected for multiple comparisons using the Holm method, revealed significant differences between all manipulations, except for the contrast between baseline and the enhancing condition. These findings indicated that the distractor and dual-task condition resulted in lower self-reported attention, but participants did not report higher attention in the enhancing condition (Table 1).

Similar results were observed for the performance test scores (Figure 3b). A generalized linear mixed-effects model confirmed a significant effect of manipulation on performance [$\chi^2(3) = 262.32$, $p < .001$]. Contrast analyses with Holm correction for multiple comparisons revealed significant differences between all manipulations in the expected direction (Table 1). These findings indicated that distractors and dual-tasks resulted in significantly lower performance than baseline, while the enhancing condition showed higher performance. Although participants did not report higher attention under the enhancing condition, these results suggest that they did remember more from these periods.

Figure 3

Behavioural measures and NET for each condition





The left panel above (3a) shows the self-reported attention scores; the right panel above (3b) shows the proportion correct answers on the performance test; the left panel below (3c) shows NET during the manipulation windows versus during the baseline data across all blocks; the right panel below (3d) shows NET for an entire block for each condition, with block 4 as baseline, blocks 2 and 6 as the enhancing condition, blocks 3 and 7 as the dual task condition and block 5 as the distracting condition. Each dot in these plots represents the mean per condition of a single participant.

Table 1

Contrast analyses for behavioural measures

		Estimate	SE	z	p_{holm}
		Self-reported attention scores			
Baseline	Enhancing	0.13	0.19	0.69	.491
	Distracting	1.21	0.23	5.23	< .001***
	Dual-task	2.06	0.19	10.94	< .001***
Enhancing	Distracting	1.08	0.23	4.67	< .001***
	Dual-task	1.93	0.19	10.25	< .001***
Distracting	Dual-task	0.86	0.23	3.70	< .001***
		Performance test			
Baseline	Enhancing	-0.09	0.04	-2.26	.024*
	Distracting	0.26	0.04	6.46	< .001***
	Dual-task	0.55	0.03	17.43	< .001***
Enhancing	Distracting	0.35	0.05	7.54	< .001***
	Dual-task	0.64	0.04	16.69	< .001***
Distracting	Dual-task	0.29	0.04	7.24	< .001***

**p* < .05. ***p* < .01. ****p* < .001.

3.2. NET

Figure 3c shows the boxplots for NET measured during the manipulation windows, separated by condition. The linear mixed-effects model analysis revealed a significant effect of manipulation on NET [$F(3,267) = 41.99$, $p < .001$]. The results of the contrast analyses with Holm correction for multiple comparisons are presented in Table 2. NET was significantly lower for the dual-task condition compared to all other conditions and for the distracting condition compared to the enhancing condition. No significant differences were found between baseline condition and both the enhancing and distracting conditions.

Next, we examined NET measured over an entire block, which consisted of a mix of data where attention was manipulated and data where no manipulation occurred, for each condition. These boxplots are shown in Figure 3d. The linear mixed-effect model analysis again found a significant effect of manipulation on NET [$F(3,147) = 18.16$, $p < .001$]. In the contrast analyses using Holm correction for multiple comparisons (Table 2), only NET in the dual-task condition was significantly lower compared to all other conditions. No other significant differences were observed.

Table 2

Contrast analyses for NET

		Estimate	SE	z	p_{holm}
		Manipulation windows			
Baseline	Enhancing	-0.01	0.01	-0.98	.327
	Distracting	0.02	0.01	1.43	.303
	Dual-task	0.07	0.07	6.58	<.001***
Enhancing	Distracting	0.03	0.03	3.41	.002**
	Dual-task	0.08	0.08	10.70	<.001***
Distracting	Dual-task	0.05	0.05	7.28	<.001***
		Entire block			
Baseline	Enhancing	$-6{,}72\text{x}10^{-5}$	0.01	-0.01	.991
	Distracting	0.01	0.01	1.62	.211
	Dual-task	0.03	0.01	5.49	<.001***
Enhancing	Distracting	0.01	0.01	1.88	.180
	Dual-task	0.03	0.01	6.74	<.001***
Distracting	Dual-task	0.02	0.01	3.62	.001**

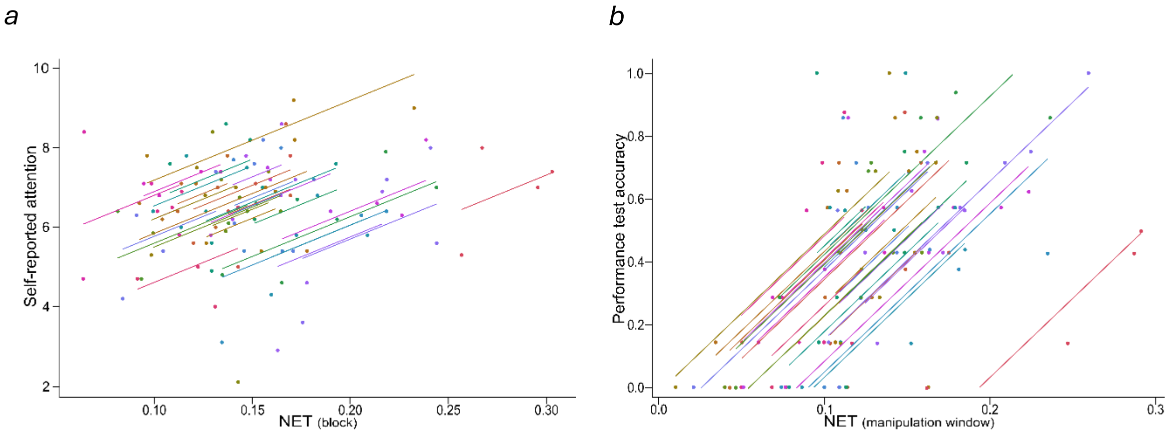
*$p < .05$. **$p < .01$. ***$p < .001$.

3.3. Associations between behavioural and EEG measures

We further examined the association between the behavioural data and NET via a repeated measures correlation analysis. The findings reveal a moderate repeated measure correlation between NET measured over an entire block and self-reported attention scores ($r_{rm} = .41$; $p < .001$). This means that an increase in NET, which reflects higher attention, is associated with higher self-reported attention, as illustrated in Figure 4a. A strong correlation was observed between NET measured exactly during the manipulation window and performance on the corresponding items ($r_{rm} = .69$; $p < .001$; Figure 4b).

Figure 4

Correlations between behavioural data and EEG data



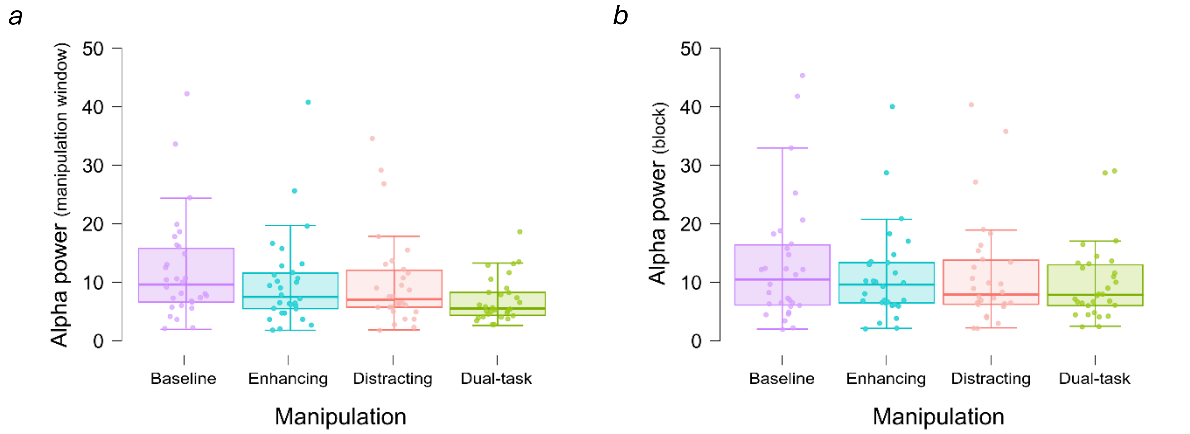
The left panel (4a) illustrates the correlations between self-reported attention and NET measured over an entire block. The right panel (4b) illustrates the correlation between the performance test accuracy and NET during the manipulation windows. The different colours represent individual participants, with each participant having four dots that indicate their average scores for each condition. The slope of the lines reflects the overall correlation between the behavioural an the EEG data.

3.4. Alpha power

We examined alpha power during the manipulation windows (Figure 5a) and alpha power for an entire block (Figure 5b). For the alpha power during the manipulation window, a significant effect of manipulation was found [$F(3,267) = 19.40$, $p < .001$]. Contrast analyses using Holm correction for multiple comparison revealed that alpha power was significantly higher in baseline compared to the other conditions. Alpha power was lower in the dual task condition, as compared to the enhancing and distraction conditions, while the latter two did not differ (Table 3). For the alpha power across the entire block, there was a significant effect of manipulation [$F(3,147) = 8.58$, $p <$ .001], but only the difference between baseline and the dual-task condition and between baseline and the enhancing condition remained significant while controlling for multiple comparisons using the Holm correction (Table 3).

Figure 5

Alpha power



The left panel (5a) illustrates the alpha power per condition during the manipulation window. The right panel (5b) illustrates the alpha power per condition for an entire block. Each dot in these plots represents the mean per condition of a single participant.

Table 3

Contrast analyses for alpha power

		Estimate	SE	z	p_{holm}
		Manipulation window			
Baseline	Enhancing	2.15	0.77	2.78	.016*
	Distracting	2.04	0.77	2.64	.017*
	Dual-task	5.05	0.77	6.54	<.001***
Enhancing	Distracting	- 0.11	0.55	- 0.20	.841
	Dual-task	2.90	0.55	5.31	<.001***
Distracting	Dual-task	3.01	0.55	5.51	<.001***
		Entire block			
Baseline	Enhancing	2.12	0.67	3.18	.007**
	Distracting	1.81	0.77	2.36	.074
	Dual-task	3.35	0.67	5.03	<.001***
Enhancing	Distracting	-0.31	0.67	-0.46	.643
	Dual-task	1.23	0.55	2.26	.074
Distracting	Dual-task	1.54	0.67	2.31	.074

*p < .05. **p < .01. ***p < .001.

4. Discussion

This study examined the use of neural envelope tracking (NET) in an educational setting. The NET reflects the strength of the relation between the EEG-based reconstructed speech stimulus (envelope) and the actual speech stimulus (envelope), as a measure of auditory attention (Biesmans et al., 2017; Roebben et al., 2024; Vanthornhout et al., 2019). To validate this method, we introduced different manipulations, such as distractors or attention enhancers, within a lecture. The findings indicated that NET, when measured during the attention manipulation window, successfully detected most of the differences between conditions in which attention was manipulated. When NET was measured over an entire block of 10 minutes (with only a few attention manipulations during those 10 minutes), only the differences between the dual-task and the other conditions were strong enough. To further validate NET, we included a performance test and a self-reported attention questionnaire. Both methods effectively distinguished between nearly all conditions, thereby confirming that our manipulations effectively modulated the participants' attention over time. Performance and self-reported attention were lower in the distraction condition and the dual-task condition compared to baseline. The enhancing condition led to improved performance, but did not result in higher self-reported attention. NET measured during the manipulation window showed a strong relation with performance test scores, while NET measured over an entire block demonstrated a weaker yet still statistically significant association with self-reported attention. All these findings support the validity of NET as a meaningful indicator of auditory attention in an educational setting.

One of the key findings of our study is that NET, whether measured during the manipulation window or over an entire block, was consistently lower when participants were engaged in and focused on the dual tasks compared to all other conditions. This finding suggests that NET is a reliable marker for detecting when attention is diverted from the lecture. A similar effect was observed by Roebben et al. (2024) and Vanthornhout et al. (2019), who investigated NET by including dual tasks into their paradigms and also observed reduced NET in these conditions. However, while both studies used children stories as stimuli, our findings extend beyond this by applying this method in a novel educational context, demonstrating its usefulness for measuring auditory attention to an educational stimulus, that was much longer in duration.

For the distracting condition, in which participants were distracted by background noise or a silent video, there was only a significant difference with the enhancing condition when NET was measured during the manipulation window. No significant difference was observed between baseline and the distracting condition. However, because NET was slightly higher in the enhancing condition and slightly lower in the distracting condition compared to baseline when measured during the manipulation window, the difference between the enhancing and distracting condition was significant. This finding aligns with the observations reported by Levy, Korisky, et al., (2025), who examined the impact of background noise on attention decoding and similarly found worse NET in the noise conditions.

For the enhancing condition, we only observed a significant difference with the distraction condition when NET was measured during the manipulation window. Enhancing did not increase NET enough to be significantly different from baseline. To the best of our knowledge, no previous studies have reported the effect on NET during such an attention enhancing condition. This lack of prior research makes it difficult to make comparisons between the current findings and previous work. It underscores the necessity for future studies to investigate the effect of attention enhancement on NET.

Besides NET measured during the manipulation window, we also decided to include NET measured over an entire block, to facilitate a fairer comparison with our self-reported attention questionnaire, which required participants to rate their attention over the same period. However, because this measure included both manipulated and non-manipulated data, the effects of our manipulations were likely diluted, as no differences were observed between baseline, enhancing and distracting conditions. Averaging the EEG signal over an entire block reduced its precision, although it is one of the key strengths of EEG to track cognitive processes during learning rather than retrospectively. However, this step was important to link it with the behavioural data of the self-reported questionnaire. Future studies can use NET for continuous monitoring of attention in the classroom and for studying learning as it happens which will be more sensitive to changes in attention similarly as our NET measured during the manipulations window.

The performance test revealed significant differences between all conditions in the expected directions, consistent with our hypotheses confirming the effect of our manipulations. This aligns with previous studies investigating the impact of manipulations on learning outcomes, for instance, Levy, Korisky, et al., (2025), found slightly lower performance in the noise conditions compared to the quiet condition, although this result was not significant. More specifically, Zeamer & Fox Tree (2013) examined how different types of auditory distractions impact test performance. They found that unexpected or unusual background sounds were more disruptive than other familiar environmental noises, such as a murmuring audience. Our findings support this observation. Despite our efforts to use naturalistic distractors, the auditory distractors in this study were unexpected and clearly not part of the original audio, making them more distracting to participants. Similarly, research has shown that enhancing attention through quizzes can have a beneficial effect on learning (Schacter & Szpunar, 2015). This was also observed in the present study, with participants achieving higher scores on the performance test for parts where attention was enhanced.

For our measures of self-reported attention, we observed similar results as those attained in our performance test, with significant differences between all manipulations. However, we did not find a significant effect of enhancing attention on the self-reports, despite the higher performance. Again, this discrepancy may be because the self-reported attention scores reflected attention during an entire block, including only a small part of enhanced attention, whereas the performance test scores were based on information from the manipulation windows. For the distraction and dual-task condition, the self-reported attention was significantly lower as compared to baseline. This aligns with the findings from Blasiman et al. (2018) and Risko et al. (2013), who reported lower self-reported attention in the conditions where participants were asked to perform other tasks while still focusing on the lecture, compared to a condition where full attention could be paid to the stimulus.

To further validate NET as a novel measure of attention in educational research, we examined the association between NET and learning outcomes with the performance test and between NET and self-reported attention with a questionnaire. We observed a strong correlation between NET measured during the manipulation window and performance test scores which further supports the validity of NET as a measure of attention. Other studies investigating the association between attention, measured using brain-to-brain synchrony or alpha power, and learning outcomes in an educational context have found similar associations (Cohen et al., 2018; Davidesco et al., 2023). However, we found no studies directly investigating the relation between NET and learning outcomes. While Levy, Korisky, et al., (2025) assessed the impact of background noise on both performance and NET, finding both to be (slightly) lower in the noise conditions, they did not explicitly examine the association between these two measures.

Similarly, we found an association between self-reported attention and NET measured over an entire block, although this association was somewhat weaker than the association between performance and NET measured during the manipulation window. As noted above, NET measured over an entire block spans a longer period and includes both manipulation and non-manipulation data and consequently masks important fluctuations in attention. The weaker association may also support the idea that it is difficult to accurately assess one's own attention through self-reports, particularly when it is over an extended period. Retrospective self-reports miss the sensitivity required to measure attention during learning, therefore misses valuable insights into the learning process (De Smedt, 2018; Mayer, 2017). Capturing continuous changes in attention during learning is difficult to achieve with behavioural measures, although the use of clickers, as in the study by Bunce et al. (2010), approximates this continuous measurement. However, clickers only provide information immediately after an attention lapse, making it difficult to identify exactly when attention begins to decrease. This highlights the significant potential of NET as a more precise and continuous measure of attention during learning that provides the opportunity to more accurately identify attention lapses.

Previous research explored the use of EEG and more specifically alpha power, a more traditional neural marker of individual attention, to measure attention in the classroom (Dikker et al., 2020; Grammer et al., 2021). In contrast to NET, where higher correlations indicate more attention, lower alpha power is associated with higher attention levels. We also analysed alpha power in order to assess how it compares to NET.

We found that alpha power measured during the manipulation window was significantly higher in the baseline condition compared to the other conditions, suggesting that baseline had the lowest level of attention. This may be explained by the introduction of additional stimuli in the enhancing and distracting conditions. Specifically, in the enhancing condition, a large red exclamation mark and a red square appeared around the video, while in the distracting condition an additional auditory or visual stimulus was added. The introduction of these new elements may have temporarily increased attention, though not necessarily towards the lecture itself. Instead, attention may have been captured by these distractors, resulting in lower alpha power in these conditions compared to baseline. Notable, the distracting condition showed lower alpha power than baseline, indicating higher attention, yet this was inconsistent with performance test scores and self-reported attention, which both indicated lower attention in this condition.

A similar discrepancy emerged in the dual-task condition, where alpha power was lowest, suggesting the highest attention compared to other conditions. However, both NET and behavioural measures indicated low attention levels for this condition. This discrepancy is not unexpected, as participants were not inattentive per se, but were engaged in the dual task rather than the lecture. This underscores a fundamental limitation of alpha power: while it effectively captures overall attentional engagement, it does not differentiate between attention towards the lecture and attention allocated to unrelated tasks. Whether students are listening to the teacher, writing, reading, or browsing through social media on their smartphone, alpha power may indicate heightened attention in all those cases, even when attention is not focused on the intended learning material.

This distinction is crucial in understanding why other studies successfully used alpha power to measure attention in the classroom (Dikker et al., 2020; Grammer et al., 2021). These studies examined attention across different teaching styles (Grammer et al., 2021) or times of the day (Dikker et al., 2020) assuming students were attending to the educational content in all conditions. In contrast, our study explicitly manipulated attention by introducing distractors and dual-tasks which deliberately diverted attention away from the lecture. Alpha power does not inherently distinguish between attention directed toward the instructional content and attention captured by other stimuli, making it less sensitive for distractions caused by other stimuli.

This limitation underscores the advantage of NET, which takes the source of attention into account rather than merely measuring overall attentional engagement. Furthermore, NET aligned more closely with the behavioural measures of attention, reinforcing its validity as a measure of attention. These findings suggest that NET offers a more accurate and nuanced approach in assessing attention in the classroom, offering new opportunities for future research to explore attentional dynamics in educational settings with greater precision.

4.1. Limitations and future directions

While the present study provides valuable insights into neural envelope tracking as a measure of auditory attention within an educational setting, several limitations should be considered when interpreting the findings. First, the objective was to utilise NET in a novel educational environment. We made the stimulus as educationally realistic as possible by using real lecture content, yet participants were seated alone in the lab, which is still different from a real classroom environment. We also used a 64-electrode EEG system with wet electrodes, which is difficult to use in real classroom settings or to simultaneously collect EEG data in multiple participants. Future research should therefore consider the use of user-friendly mobile EEG devices. However these systems typically have a lower number of electrodes, and tend to be more sensitive to various artifacts, leading to increased data loss, making further improvements necessary for broader implementation in educational research (Janssen et al., 2021). Despite these challenges, mobile-EEG systems would allow to use NET as a measure of attention in a real classroom environment, with all its inherent features, such as classmates, a familiar teacher, and more real-world distractions. This is important because previous studies that used brain-to-brain synchrony to measure attention have shown that this measure is influenced by factors such as social closeness between classmates (Dikker et al., 2020) and between students and teachers (Bevilacqua et al., 2019), but it remains unclear whether such factors also impact NET and this represents an avenue for future study.

Second, in the present experiment, we collected a large amount of baseline data, which was needed to train the decoder. During this baseline, the participant's attention was not manipulated, yet participants still experienced natural fluctuations in attention due to factors such as personal thoughts, fatigue, hunger, and environmental distractions. While we attempted to account for these variations by averaging data over a larger time window and including segments throughout the lecture, this approach may not fully capture a stable 'baseline attentional state', which is crucial for training the decoder. If baseline attention itself is highly variable, it may introduce noise into the decoding process and affect the reliability of NET. Future research should explore methods, such as the use of unsupervised (time-adaptive) training for such decoders (Geirnaert et al., 2022; Heintz et al., 2025), to refine baseline measurements to improve the accuracy of attention decoding, particularly for detecting subtle, internally driven fluctuations in attention.

Third, and related to the previous point, a large amount of training data is needed. The present experiment is lengthy, but the amount of available EEG data that can be analysed is limited because almost half of the EEG data is used to train the decoder. This results in a significant data loss for the actual analyses. Future research should aim to reduce the amount of training data required, which might be done by starting from universal decoders and applying unsupervised (time-adaptive) training to gradually finetune them towards a personalized decoder (Geirnaert et al., 2022; Heintz et al., 2025). This is necessary to decrease the amount of training data and hopefully provide opportunities to give real-time feedback on students' attention in the future.

Fourth, the sample primarily consisted of participants who had either obtained or were pursuing a higher education degree. Given their likely experience with sustaining attention for extended periods and engaging with online lectures, this may have influenced the results. To strengthen the validation of NET, future studies should examine its sensitivity in more diverse populations, including individuals with different educational backgrounds, age groups, and attentional profiles (e.g., typically developing individuals vs. those with attentional difficulties). Such studies would help to determine whether NET can accurately capture attentional differences between groups and assess its alignment with self-reported attentional states in different populations.

Lastly, while the present findings provide evidence for the effectiveness of this method in measuring attention during lecture-based teaching, its applicability to other instructional formats remains uncertain. For teaching methods involving group work or discussions where multiple speakers interact, it remains unclear whether this method can reliably capture attention. While previous studies have used this method to identify the source someone is paying attention to in a competing speakers paradigm (Mirkovic et al., 2015; O'Sullivan et al., 2015), further research is necessary to assess its effectiveness in absolute attention decoding in more interactive learning environments.

4.2. Conclusion

The current study provides evidence to support the use of neural speech tracking as a valid measure of auditory attention in educational settings. By incorporating the stimulus, that is the instructor's voice, into the calculation of the attention measure, neural envelope tracking provides a more accurate assessment of attention to the teacher compared to the commonly used alpha power measure. Future research should optimize this technique by focusing on more user-friendly EEG devices, such as mobile EEG, the use of self-adapting decoders, and by studying its applications in more diverse populations and educational settings. This will allow us to leverage the possibilities of NET into real-world educational studies.

References

Bakdash, J. Z., & Marusich, L. R. (2017). Repeated Measures Correlation. *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.00456

Bakdash J, Marusich L (2024*). _rmcorr: Repeated Measures Correlation_* (Version 0.7.0) [Computer software]. https://CRAN.R-project.org/package=rmcorr

Bevilacqua, D., Davidesco, I., Wan, L., Chaloner, K., Rowland, J., Ding, M., Poeppel, D., & Dikker, S. (2019). Brain-to-brain synchrony and learning outcomes vary by student–teacher dynamics: Evidence from a real-world classroom electroencephalography etudy. *Journal of Cognitive Neuroscience*, *31*(3), 401–411. https://doi.org/10.1162/jocn_a_01274

Biesmans, W., Das, N., Francart, T., & Bertrand, A. (2017). Auditory-inspired speech envelope extraction methods for improved EEG-based auditory attention detection in a cocktail party scenario. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *25*(5), 402–412. https://doi.org/10.1109/TNSRE.2016.2571900

Biosemi. (n.d.). *Active2 and Active 2.5: 280-channel, DC amplifier, 24-bit resolution, biopotential measurement system with active electrodes.* https://www.biosemi.com/Products_ActiveTwo.htm

Blasiman, R. N., Larabee, D., & Fabry, D. (2018). Distracted students: A comparison of multiple types of distractions on learning in online lectures. *Scholarship of Teaching and Learning in Psychology*, *4*(4), 222–230. https://doi.org/10.1037/stl0000122

Booth, B. M., Bosch, N., & D'Mello, S. K. (2023). Engagement detection and its applications in learning: A tutorial and selective review. *Proceedings of the IEEE*, *111*(10), 1398–1422. https://doi.org/10.1109/JPROC.2023.3309560

Boudewyn, M. A., & Carter, C. S. (2018). I must have missed that: Alpha-band oscillations track attention to spoken language. *Neuropsychologia*, *117*, 148–155. https://doi.org/10.1016/j.neuropsychologia.2018.05.024

Bradbury, N. A. (2016). Attention span during lectures: 8 seconds, 10 minutes, or more? *Advances in Physiology Education*, *40*(4), 509–513. https://doi.org/10.1152/advan.00109.2016

Bunce, D. M., Flens, E. A., & Neiles, K. Y. (2010). How long can students pay attention in class? A study of student attention decline using clickers. *Journal of Chemical Education*, *87*(12), 1438–1443. https://doi.org/10.1021/ed100409p

Cohen, S. S., Madsen, J., Touchan, G., Robles, D., Lima, S. F. A., Henin, S., & Parra, L. C. (2018). Neural engagement with online educational videos predicts learning performance for individual students. *Neurobiology of Learning and Memory*, *155*, 60–64. https://doi.org/10.1016/j.nlm.2018.06.011

Davidesco, I. (2020). Brain-to-brain synchrony in the STEM classroom. *CBE* Life Sciences Education, *19*(3). Article es8. https://doi.org/10.1187/cbe.19-11-0258

Davidesco, I., Laurent, E., Valk, H., West, T., Milne, C., Poeppel, D., & Dikker, S. (2023). The temporal dynamics of brain-to-brain synchrony between students and teachers predict learning outcomes. *Psychological Science*, *34*(5), 633–643. https://doi.org/10.1177/09567976231163872

De Smedt, B. (2018). Applications of cognitive neuroscience in educational research. In *Oxford Research Encyclopedia of Education*. https://doi.org/10.1093/acrefore/9780190264093.013.69

Dikker, S., Haegens, S., Bevilacqua, D., Davidesco, I., Wan, L., Kaggen, L., McClintock, J., Chaloner, K., Ding, M., West, T., & Poeppel, D. (2020). Morning brain: Real-world neural evidence that high school class times matter. *Social Cognitive and Affective Neuroscience*, *15*(11), 1193–1202. https://doi.org/10.1093/scan/nsaa142

Dikker, S., Wan, L., Davidesco, I., Kaggen, L., Oostrik, M., McClintock, J., Rowland, J., Michalareas, G., Bavel, J. J. V., Ding, M., & Poeppel, D. (2017). Brain-to-brain synchrony tracks real-world dynamic group interactions in the classroom. *Current Biology*, *27*(9), 1375–1380. https://doi.org/10.1016/j.cub.2017.04.002

Dmochowski, J. P., Sajda, P., Dias, J., & Parra, L. C. (2012). Correlated components of ongoing EEG point to emotionally laden attention: A possible marker of engagement? *Frontiers in Human Neuroscience*, *6*, Article 112. https://doi.org/10.3389/fnhum.2012.00112

Foxe, J., & Snyder, A. (2011). The role of alpha-band brain oscillations as a sensory suppression mechanism during selective attention. *Frontiers in Psychology*, *2,* Article 154. https://doi.org/10.3389/fpsyg.2011.00154

Geirnaert, S., Francart, T., & Bertrand, A. (2022). Time-adaptive unsupervised auditory attention decoding using EEG-based stimulus reconstruction. *IEEE Journal of Biomedical and Health Informatics*, *26*(8), 3767–3778. https://doi.org/10.1109/JBHI.2022.3162760

Grammer, J. K., Xu, K., & Lenartowicz, A. (2021). Effects of context on the neural correlates of attention in a college classroom. *Npj Science of Learning*, *6*, Article 15. https://doi.org/10.1038/s41539-021-00094-8

Heintz, N., Francart, T., & Bertrand, A. (2025). *Unsupervised EEG-based decoding of absolute attention with canonical correlation analysis*. ArXiv. https://doi.org/10.48550/arXiv.2504.17724

Hlas, A. C., Neyers, K., & Molitor, S. (2019). Measuring student attention in the second language classroom. *Language Teaching Research*, *23*(1), 107–125. https://doi.org/10.1177/1362168817713766

Hobbiss, M. H., & Lavie, N. (2024). Sustained selective attention in adolescence: Cognitive development and predictors of distractibility at school. *Journal of Experimental Child Psychology*, *238*, Article 105784. https://doi.org/10.1016/j.jecp.2023.105784

Howard-Jones, P. (2010). *Introducing neuroeducational research: Neuroscience, education and the brain from contexts to practice*. Routledge.

Janssen, T. W. P., Grammer, J. K., Bleichner, M. G., Bulgarelli, C., Davidesco, I., Dikker, S., Jasińska, K. K., Siugzdaite, R., Vassena, E., Vatakis, A., Zion-Golumbic, E., & van Atteveldt, N. (2021). Opportunities and limitations of mobile neuroimaging technologies in educational neuroscience. *Mind, Brain, and Education*, *15*(4), 354–370. https://doi.org/10.1111/mbe.12302

Jarrett, C. (2015). *Great myths of the brain*. Wiley Blackwell.

JASP Team (2024). *JASP* (Version 0.19.0) [Computer software]. https://jasp-stats.org/

Keller, A. S., Davidesco, I., & Tanner, K. D. (2020). Attention matters: How orchestrating attention may relate to classroom learning. *CBE Life Sciences Education*, *19*(3), Article fe5. https://doi.org/10.1187/cbe.20-05-0106

Ki, J. J., Kelly, S. P., & Parra, L. C. (2016). Attention strongly modulates reliability of neural responses to naturalistic narrative stimuli. *The Journal of Neuroscience*, *36*(10), 3092–3101. https://doi.org/10.1523/JNEUROSCI.2942-15.2016

Levy, O., Hackmon, S. L., Zvilichovsky, Y., Korisky, A., Bidet-Caulet, A., Schweitzer, J. B., & Golumbic, E. Z. (2025). *Selective attention and sensitivity to auditory disturbances in a virtually-real Classroom: Comparison of adults with and without AD(H)D*. bioRxiv. https://doi.org/10.1101/2024.04.17.590012

Levy, O., Korisky, A., Zvilichovsky, Y., & Zion Golumbic, E. (2025). The neurophysiological costs of learning in a noisy classroom: An ecological virtual reality study. *Journal of Cognitive Neuroscience*, *37*(2), 300–316. https://doi.org/10.1162/jocn_a_02249

Li, T., Geirnaert, S., Van den Broek, D., Bellon, E., De Smedt, B., & Bertrand, A. (2025). *Temporal variation in the acoustic dynamic range is a confounding factor in EEG-based tracking of absolute auditory attention to speech*. Proc. International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Denmark.

Malik, A. S., & Amin, H. U. (2017). *Designing EEG experiments for studying the brain*. Academic Press. https://doi.org/10.1016/B978-0-12-811140-6.00001-1

MATLAB. (2010). Version 9.11 (2021b) [Computer software]. Natick Massachusetts: The Mathworks Inc.

Matsuo, M., Higuchi, T., Ichibakase, T., Suyama, H., Takahara, R., & Nakamura, M. (2024). Differences in electroencephalography power levels between poor and good performance in attentional tasks. *Brain Sciences*, *14*(6), Article 6. https://doi.org/10.3390/brainsci14060527

Mayer, R. E. (2017). How can brain research inform academic learning and instruction? *Educational Psychology Review*, *29*(4), 835–846. https://doi.org/10.1007/s10648-016-9391-1

Mirkovic, B., Debener, S., Jaeger, M., & De Vos, M. (2015). Decoding the attended speech stream with multi-channel EEG: Implications for online, daily-life applications. *Journal of Neural Engineering*, *12*(4), Article 046007. https://doi.org/10.1088/1741-2560/12/4/046007

O'Connell, R. G., Dockree, P. M., Robertson, I. H., Bellgrove, M. A., Foxe, J. J., & Kelly, S. P. (2009). Uncovering the neural signature of lapsing attention: Electrophysiological signals predict errors up to 20 s before they occur. *Journal of Neuroscience*, *29*(26), 8604–8611. https://doi.org/10.1523/JNEUROSCI.5967-08.2009

Olney, A. M., Risko, E. F., D'Mello, S. K., & Graesser, A. C. (2015). Attention in educational contexts: The role of the learning task in guiding attention. In J. Fawcett, E. Risko, & A. Kingstone (Eds.) *The handbook of attention* (pp. 623-641). The MIT Press. https://doi.org/10.7551/mitpress/10033.003.0030

O'Sullivan, J. A., Power, A. J., Mesgarani, N., Rajaram, S., Foxe, J. J., Shinn-Cunningham, B. G., Slaney, M., Shamma, S. A., & Lalor, E. C. (2015). Attentional selection in a cocktail party environment can be decoded from single-trial EEG. *Cerebral Cortex*, *25*(7), 1697–1706. https://doi.org/10.1093/cercor/bht355

Poulsen, A. T., Kamronn, S., Dmochowski, J., Parra, L. C., & Hansen, L. K. (2017). EEG in the classroom: Synchronised neural recordings during video presentation. *Scientific Reports*, *7*, Article 43916. https://doi.org/10.1038/srep43916

Rapport, M. D., Kofler, M. J., Alderson, R. M., Timko, T. M., & DuPaul, G. J. (2009). Variability of attention processes in ADHD: Observations from the classroom. *Journal of Attention Disorders*, *12*(6), 563–573. https://doi.org/10.1177/1087054708322990

Risko, E. F., Buchanan, D., Medimorec, S., & Kingstone, A. (2013). Everyday attention: Mind wandering and computer use during lectures. *Computers & Education*, *68*, 275–283. https://doi.org/10.1016/j.compedu.2013.05.001

R Core Team (2024). _*R: A Language and Environment for Statistical Computing*_ (Version 4.4.1) [Computer software]. https://www.R-project.org/

Roebben, A., Heintz, N., Geirnaert, S., Francart, T., & Bertrand, A. (2024). 'Are you even listening?': EEG-based decoding of absolute auditory attention to natural speech. *Journal of Neural Engineering*, *21*(3), Article 036046. https://doi.org/10.1088/1741-2552/ad5403

Schacter, D. L., & Szpunar, K. K. (2015). Enhancing attention and memory during video-recorded lectures. *Scholarship of Teaching and Learning in Psychology*, *1*(1), 60–71. https://doi.org/10.1037/stl0000011

Straetmans, L., Adiloglu, K., & Debener, S. (2024). Neural speech tracking and auditory attention decoding in everyday life. *Frontiers in Human Neuroscience*, *18*. https://doi.org/10.3389/fnhum.2024.1483024

Unsworth, N., & McMillan, B. D. (2017). Attentional disengagements in educational contexts: A diary investigation of everyday mind-wandering and distraction. *Cognitive Research: Principles and Implications*, *2*(32), 1-20. https://doi.org/10.1186/s41235-017-0070-7 van Atteveldt, N., van Kesteren, M. T. R., Braams, B., & Krabbendam, L. (2018). Neuroimaging of learning and development: Improving ecological validity. *Frontline Learning Research*, *6*(3), 186–203. https://doi.org/10.14786/flr.v6i3.366

Vanthornhout, J., Decruy, L., & Francart, T. (2019). Effect of task and attention on neural tracking of speech. *Frontiers in Neuroscience*, *13*. Article 977. https://doi.org/10.3389/fnins.2019.00977

Zeamer, C., & Fox Tree, J. E. (2013). The process of auditory distraction: Disrupted attention and impaired recall in a simulated lecture environment. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(5), 1463–1472. https://doi.org/10.1037/a0032190