Neurophysiological Patterns of Attention and Distraction during Realistic Virtual-Reality Classroom Learning in Adults with and without ADHD

Orel Levy[1], Shirley Libman Hackmon[1], Yair Zvilichovsky[1], Adi Korisky[1], Aurelie Bidet-Caulet[2], Julie B. Schweitzer[3], *Elana Zion Golumbic[1]

[1]The Gonda Brain Research Center, Bar Ilan University, Ramat Gan, Israel

[2] Aix Marseille Univ, Inserm, INS, Inst Neurosci Syst, Marseille, France

[3]Department of Psychiatry and Behavioral Sciences, and [4] MIND Institute, University of California, Davis, Sacramento, CA U.S.A.

*corresponding author

Prof. Elana Zion Golumbic

email: elana.zion-golumbic@biu.ac.il

The Gonda Brain Research Center

Bar Ilan University

Ramat Gan, Israel

Abstract

Many people, and particularly those diagnosed with ADHD, report difficulties maintaining attention and proneness to distraction during classroom learning. However, the behavioral, neural and physiological basis of attention in realistic learning contexts is not well understood, since current clinical and scientific tools used for evaluating and quantifying the constructs of "distractibility" and "inattention", are removed from the real-life experience in organic classrooms. Here we introduce a novel Virtual Reality (VR) platform for studying students' brain activity and physiological responses as they immerse in realistic frontal classroom learning.

Using this approach, we studied whether adults with and without ADHD (N=49) exhibit differences in neurophysiological metrics associated with sustained attention, such as speech-tracking of the teacher's voice, power of alpha-oscillations and levels of arousal, as well as responses to potential disturbances by background sound-events in the classroom.

Under these ecological conditions, we find that adults with ADHD exhibit higher auditory neural response to background sounds relative to their control-peers, which also contributed to explaining variance in the severity of ADHD symptoms, together with higher power of alpha-oscillations and more frequent gaze-shifts around the classroom. These results are in-line with higher sensitivity to irrelevant stimuli in the environment and increased mind-wandering/boredom. At the same time, both groups exhibited similar learning outcomes and showed similar neural tracking of the teacher's speech. This suggests that in this context, attention may not operate as a zero-sum game and that allocating some resources to irrelevant stimuli does not always detract from performing the task at hand.

Given the dire need for more objective, dimensional and ecologically-valid measures of attention and its real-life deficits, this work provides new insights into the neurophysiological manifestations of attention and distraction experienced in real-life contexts, while challenging some prevalent notions regarding the nature of attentional challenges experienced by those with ADHD.

Introduction

Many people find maintaining attention to a frontal classroom lecture difficult. It requires listening, processing and comprehending the teacher's speech over a prolonged period of time, while avoiding distraction from both irrelevant background sounds and internal mind-wandering (Smallwood et al., 2007; Szpunar et al., 2013; Thomson et al., 2015; Esterman and Rothlein, 2019). This task is supposedly even more difficult for individuals diagnosed with Attention-Deficit Hyperactivity Disorder (ADHD), which is often characterized by increase proneness to distraction and mind-wandering and poorer sustained attention (Avisar and Shalev, 2011; Berger and Cassuto, 2014; Merrill et al., 2022). Unsurprisingly perhaps, individuals are most likely to seek clinical help as a result of attentional difficulties experienced in school contexts (Litner, 2003), and there is a tight link between attention performance and learning outcomes (Loe and Feldman, 2007; Gray et al., 2017).

However, current clinical and scientific tools used for evaluating and quantifying the constructs of "distractibility" and "inattention," are greatly removed from the real-life experience in organic classrooms and other ecological contexts. Commonly used approaches either rely on subjective self-report (e.g., rating scales), or are based on highly artificial computerized tests, such as Continuous Performance Tests (CPT). Unfortunately, in recent years there is a growing understanding that these tools lack sufficient sensitivity and specificity to reliably capture the type and severity of real-life attention challenges that people face outside the lab or the clinic (Rabiner et al., 2010; Narad et al., 2015; Hall et al., 2016; Murray et al., 2018; Mulraney et al., 2022; Hobbiss and Lavie, 2024).

In attempt to create laboratory-based tasks that are more ecologically valid, and better capture the type of attentional challenges faced in real-life, recent years have seen increased use of Virtual Reality (VR) in cognitive research (Parsons, 2015; Seesjärvi et al., 2022). In particular, VR classrooms have been used to test different aspects of cognitive performance and distraction by ecologically valid stimuli. For example, it has been suggested that CPT tasks administered in VR-Classroom environments show greater sensitivity for ADHD classification relative to standard evaluations and can provide more refined measures of attention and distractibility (Rizzo et al., 2007; Neguț et al., 2017; Coleman et al., 2019; Parsons et al., 2019; Stokes et al., 2022). However, to date, most VR-classroom studies still use relatively non-ecological tasks, and rely primarily on indirect behavioral outcomes to assess aspects of attentional performance. To improve the utility

of VR classrooms for understanding attention and distraction in real-life classroom learning, these platforms need to use genuine academic tasks and stimuli, to more faithfully simulate the classroom experience. Moreover, there is need to incorporate more objective metrics related to attention, such as measurements of neural activity and physiological responses, in order to obtain a more comprehensive picture of how listeners actually respond to stimuli and process the content of a lesson in a VR classroom. Here we address this challenge by introducing a highly realistic multimodal VR classroom platform, that simulates participating in a frontal lesson by an avatar-teacher with occasional disturbances from ecological sound-events. Our novel platform also integrates measurements of neurophysiological responses, including neural activity (using electroencephalography; EEG), eye-gaze and arousal levels, as reflected in changes in skin-conductance. These multi-level recordings allow us to investigate which neurophysiological metrics distinguish between students with and without a diagnosis of ADHD when they engage in realistic classroom learning, and which metrics reliably predict the severity of ADHD-symptoms.

We focus on several specific neural and physiological metrics, that have been associated with aspects of attention, distraction and information processing in more traditional attention research, and test their generalizability to this ecological context. These include: (1) Neural speech-tracking of the teacher's lesson, which represent its sensory encoding and potentially linguistic processing, and is known to be reduced under conditions of inattention or distraction (Ding and Simon, 2012a; Zion Golumbic et al., 2013; Huang and Elhilali, 2020; Holtze et al., 2021; Kaufman and Zion-Golumbic, 2023). Accordingly, this metric might be expected to be reduced in individuals with ADHD and/or in conditions that contain external disturbances. (2) The power of alpha and beta oscillations, which are often associated with increased mind-wandering or boredom (Clarke et al., 2001; Boudewyn and Carter, 2018; Jin et al., 2019), and some have suggested may be altered in individuals with ADHD (although use of spectral signatures as 'biomarkers for ADHD is highly controversial; (Gloss et al., 2016). (3) Continuous levels of arousal, as measured by skin-conductance (SC), which some propose are either heightened or reduced in individuals with ADHD relative to their control peers (Sergeant, 2000; Bellato et al., 2020). (4) Neural Event-Related Potentials (ERPs) and transient changes in SC following unexpected sound-events in the background of the classroom, which reflect aspects of exogenous capture of attention, increases in arousal and potentially distraction by salient irrelevant stimuli (Posner, 1980; Bidet-Caulet et al., 2015; Masson and Bidet-Caulet, 2019). These too would be expected to be heightened in

individuals with ADHD, reflecting their higher susceptibility to external distraction (Gumenyuk et al., 2005). (5) Eye-gaze patterns, which are commonly used as a proxy for determining the focus overt attention and detecting attention-shifts (Grosbras et al., 2005; Schomaker et al., 2017). Recent VR classroom suggest that children with ADHD tend to perform more gaze-shifts, particularly after distracting events, and take longer to re-focus on task-relevant portions of the scene, relative to control peers (Mauriello et al., 2022; Stokes et al., 2022; Selaskowski et al., 2023).

This novel ecological approach and rich neurophysiological dataset allowed us to test specific hypothesis-driven questions detailed above regarding differences between individuals with and without a diagnosis of ADHD. We also conducted exploratory analyses to better understand the relationship between different neurophysiological measures and their potential utility as objective markers for attention and distraction, under these highly realistic conditions.

Methods

Participants

Fifty-four participants initially participated in the study; however, the data from five participants were excluded due to technical issues or their voluntary request to discontinue participation. The final sample consisted of forty-nine 49 participants (21 male, 28 female). Of these participants, 24 had a prior clinical diagnosis of attention deficit hyperactivity disorder (ADHD group), and 25 participants did not (control group). The two groups were matched for age, which ranged between 20 and 38 years (mean age of all participants: 26.9 ± 3.0; ADHD participants: 26.7 ± 3.7; control participants: 25.7 ± 4.4). The majority of participants were university students. Participants in the ADHD group were questioned about their medication intake regimen, and the experiment was scheduled at least 12 hours after their last dose of medication. None of the participants (in either group) reported taking any other medications on a regular basis. The experimental protocol was approved by the ethics committee at Bar-Ilan University, and all participants provided written informed consent prior to their involvement in the study and data collection procedures. Participants received financial compensation or course credit in exchange for their participation.

ASRS questionnaire

All participants completed a self-report questionnaire assessing ADHD symptoms (ASRS-v1.1; Adler et al., 2006, Hebrew version: Zohar and Konfortes, 2010), which has been shown to have

good psychometric properties high sensitivity for ADHD diagnosis (Kessler et al., 2005). Participants in the ADHD group were instructed to complete the questionnaire based on their experiences at a time when they are not taking medication. The questionnaire consists of 18 items, which are rated on a 5-point Likert scale ranging from 0 (never) to 4 (very often). ASRS scores for each participant were calculated as the total sum of marks across all items, as this has been shown to yield the highest sensitivity to ADHD diagnosis (Zohar and Konfortes, 2010). We did not try to categorize ADHD-subtypes (e.g., inattention, hyperactivity, combined) since these subtypes are less reliable in adulthood. As expected, the group diagnosed with ADHD exhibited significantly higher ASRS scores compared to the control group [$t(47) = 8.49$, $p < 0.001$; see Figure 1]. This outcome validates the group allocation and demonstrates the consistency between ADHD diagnosis and self-reported ADHD symptoms.

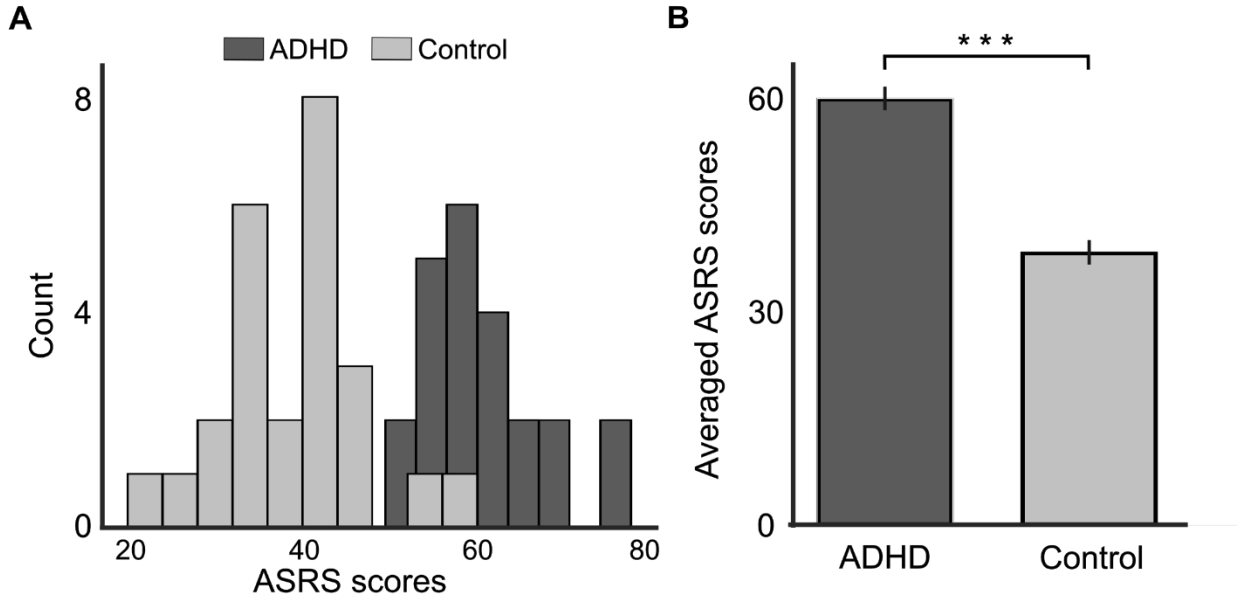

Figure 1. Histogram (A) and bar graph (B) showing the distribution of ASRS scores in the ADHD and Control groups. Error bars represent standard error of the mean (SEM). This confirms group-allocation, within the tested sample.

The VR Classroom: Design and Stimuli

The VR classroom platform was developed using the Unity game engine (unity.com; JavaScript and C# programming). Participants experienced the virtual environment through an HTC Vive ProEye VR headset that is equipped with an embedded eye-tracker (120 Hz binocular sampling

rate). Audio was delivered through in-ear headphones (Etymotic ER1-14 with disposable foam eartips).

In the VR classroom (Figure 2A), participants experienced sitting at a desk in the second row of a classroom, facing a male avatar-teacher, who was standing in front of a blackboard and a wall adorned with maps. Ten additional avatar-students occupied the remaining desks in the classroom. Although these student avatars didn't engage in verbal communication during the experiment, their bodies were animated to simulate natural sitting movements. The avatar-teacher delivered a series of 30 mini-lessons on scientific or historical topics (mean duration 42.7 sec ± 5.1). The audio for these mini-lessons was taken from existing informational podcast, recorded by a male speaker. Speech audio was presented using 3D spatial audio (implemented in Unity) to be perceive as emitting from the teacher's lips. In addition, the teacher's lip movements were animated to follow the temporal envelope of the speech, and his body motions were animated to mimic natural talking gestures and to capture appropriate emphasis and prosodic cues.

Besides the teacher's lesson, short sounds ('sound-events') were presented occasionally in most trials, from a spatial location to the right or left of the participant (randomized), as if originating from one of the students sitting beside them in class (Figure 2B). A total of 110 different sounds were used, taken from existing sound databases (freesfx.co.uk, findsounds.com, freesound.org, and the IADS-E sound stimuli database (Yang et al., 2018), from two broad categories: Non-verbal human sounds (mainly coughs and throat clearings), and Artificial sounds (phone ringtones and digitally-generated 'boing' sounds). All sound-events were normalized to the same loudness level using the software Audacity and were trimmed to a duration of 300 milliseconds.

In the VR classroom, the loudness of the teacher's speech was set to a comfortable level for each participant, as determined in a training trial. The precise intensity level cannot be measured reliably, since it is affected by the specific positioning of the foam earphone inserts in the participants' ear, but was roughly between 60-70 dB SPL. The sound-events were played at a loudness level of 50% relative to the teacher's voice (-3dB SPL).

Experimental procedure

The experiment was conducted in a sound attenuated room. After comfortably fitting the VR headset on the participant, the eye-tracker was calibrated and validated using a standard 9-point calibration procedure. Participants were then familiarized with the VR classroom scene, and

performed two training trials – one without events-sounds and one with – for training and calibration purposes.

The experiment consisted of 30 trials. In each trial, the avatar teacher delivered a different mini-lesson and participants were instructed to pay attention to its content. They were not given any specific instructions regarding the other sounds or stimuli in the classroom. After each mini-lesson, participants answered four multiple-choice questions about its content (Figure 2C). Most trials (22) contained occasional background event-sound (Events condition; with 5 of both types sound-events per trial, separated by 3-7 seconds; Figure 2B), whereas 8 trials did not, serving as a Quiet baseline condition. In total, 104 sound-events were presented throughout the experiment (52 of each type). The order of mini-lecture presentation and their allocation to the Events or Quiet condition was randomized across participants.

EEG, GSR and Eye-Gaze recordings

Neural activity was recorded using 64 active ag-agCl EEG electrodes (BioSemi) and was sampled at 1024Hz. Electrooculographic (EOG) signals were also included in this recording, measured by 3 additional electrodes located above the right eye and on the external side of both eyes.

The Skin Conductance Response (SC; Akash et al., 2018) was recorded using two passive Ni hon Kohden electrodes placed on the fingertips of the index and middle fingers of participants' non-dominant hand. These signals were also sampled by the BioSemi system, and thus synchronized to EEG data.

Eye-gaze data was recorded using the eye tracker embedded in the VR headset, which provides information about the x-y-z coordinate of the gaze in 3D space at each time-point, as well as automatic blink-detection. We defined several a-priori regions of interest (ROIs) in the 3D classroom space, which included: 'Teacher', 'Left Board', 'Right Board', 'Ceiling', 'Floor', 'Middle Window', 'Left Student', 'Right Student'. Note that the sound-events were emitted from the ROI labeled 'Right Student'. Using a combined calculation of the participants' head position and direction of the gaze vector, gaze data could be automatically classified to describe which ROI the participant looked at, at each point in time.

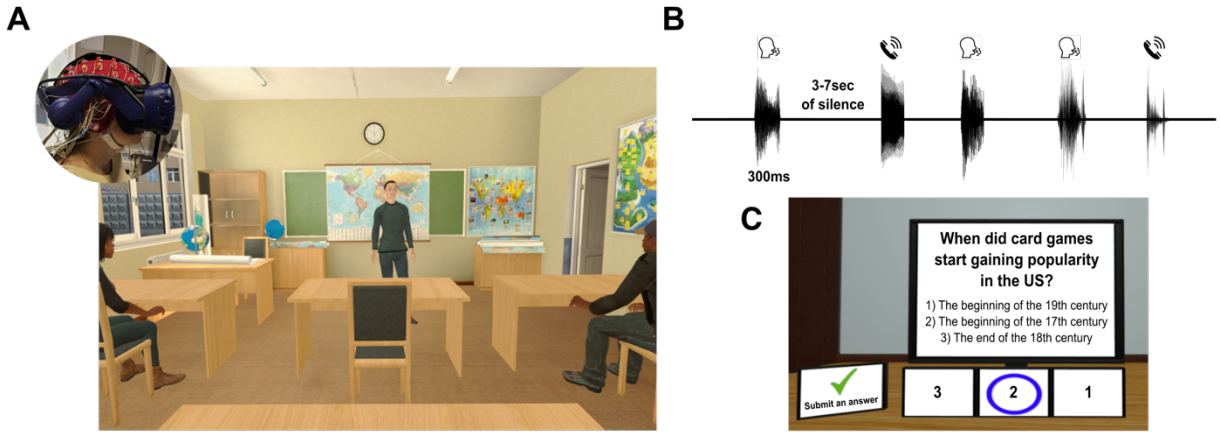

Figure 2. A. Illustration of the Virtual Classroom used in the study. Inset: a participant wearing the VR headset over the EEG cap. **B.** Example of the sequence of sound-events presented in Event-trials, which were emitted from a spatial location to the right or left of the participant. Event-trials contained 5 events, of both types (Artificial and Human non-verbal sounds), randomized across trials and separated by 3-7 seconds. **C.** Example of a multiple-choice comprehension question presented after each trial (English translation).

Behavioral data analysis

To evaluate participants' comprehension and memory retention, we calculated their accuracy on four multiple-choice questions following each trial (average correct responses). Between-group differences were assessed using unpaired t-tests, separately for the Quiet and Events conditions. These were evaluated separately (here and in all other analyses), given the imbalanced number of trials in each condition. The Bayes Factor (BF_{10}) was also calculated for each pairwise comparison, to test the level of confidence for embracing the null hypothesis (H0; BF_{10} thresholds used throughout: $BF_{10} < 1$ provide weak support of H0 and $BF_{10} < 0.3$ provide moderate/strong support of H0).

EEG data analysis

The EEG data was preprocessed using MATLAB with the FieldTrip toolbox (https://www.fieldtriptoolbox.org/). The raw EEG data was referenced to linked right and left mastoids, then band-pass filtered between 0.5 and 40 Hz, detrended and demeaned. Visual inspection was performed to identify and remove gross artifacts (excluding eye movements) and Independent Component Analysis (ICA) was then used to further remove components associated

with horizontal or vertical eye movements and heartbeats. Remaining noisy electrodes, containing extensive high-frequency activity or DC drifts, were interpolated using adjacent electrodes, either on the entire dataset or on a per-trial basis, as needed.

Three types of analyses were applied to the clean EEG data: 1) Speech tracking analysis of the neural response to teacher's speech, 2) Event related potentials (ERPs) analysis capturing the time-locked neural response to the Sound-events and 3) Spectral analysis, looking specifically at ongoing power in two canonical frequency bands previously associated with attention: the Alpha (6-15 Hz), and Beta (13–30 Hz) ranges.

Speech Tracking response (TRF)

The clean EEG data was segmented into trials from the onset the teacher's speech until the end of the trial. For this analysis, EEG data was further bandpass filtered between 0.8 and 20 Hz and downsampled to 100 Hz for computational efficiency. To estimate the neural response to the speech we performed speech-tracking analysis which estimates a linear Temporal Response Function (TRF) that best describes the relationship between features of the stimulus (S) and the neural response (R) in a given trial. Speech tracking analysis was conducted using the mTRF MATLAB toolbox (Crosse et al., 2016), using both an encoding and decoding approach. In both the Quiet and Events condition, the broadband speech envelope was used as S, derived by filtering the speech through a narrow-band cochlear filterbank, extracting the narrowband envelopes using a Hilbert transform, and averaging across bands. In the Events condition, we additionally used the time-course of the sound-events as a second regressor (S2) in a multivariate TRF model, so that it fully represents the soundscape heard by participants. Both regressors were downsampled to 100 Hz to match the EEG data and normalized using z-score.

In the **encoding** approach, TRFs are estimated for each EEG channel, with lags between the S and the R ranging from -150ms (baseline) to 450ms. The mTRF toolbox uses a ridge-regression approach for L2 regularization of the model, to reduce its complexity and ensure better generalization to new data. Based on previous research in by our group (Kaufman & Zion Golumbic 2023), we chose to use a constant regularization ridge parameter (λ= 100) for all participants and conditions. We used a leave-one-out cross-validation procedure to assess the model's predictive power, whereby in each iteration, all but one trials are used to train the model, and it is then applied to the left-out trial. The predictive power of the model is estimated as the

Pearson's correlation between the predicted neural responses and the actual neural responses, separately for each electrode, averages across all iterations. Statistical significance of the model's predictive power at each electrode is evaluated using a permutation test where the same procedure is repeated 100 times on shuffled data where the S from each trial is randomly paired with the R from a different trial. From each permutation we extract the highest predictive power across all electrodes, yielding a null-distribution of the maximal predictive power that could be obtained by chance. The predictive power in the real data is compared to this distribution, and electrodes with values falling in the top 5%^{tile} of the null-distribution are considered to have a significant speech tracking response. Note that by choosing the maximal value across electrodes in each permutation, this procedure also corrects for multiple comparisons.

TRFs were derived separately for each Group (ADHD vs. Control) and Condition (Quiet vs. Events), and for the Events Condition we report results of both the univariate model (using only the speech-regressor) and the multivariate model (which also includes the envelope of the sound-events). Between group-differences were assessed separately for each condition, given their different acoustic parameters and unequal number of trials. Unpaired t-tests were performed on the TRF at each-time point and electrode to identify spatio-temporal clusters where the TRFs differed significantly between the two groups (Monte-Carlo cluster-based permutation test; fieldtrip toolbox).

A complementary way to examine the neural representation of continuous speech is the **decoding** approach. In the decoding model a multidimensional transfer function is estimated using the neural data (R) from all electrodes as input in attempt to reconstruct the stimulus feature (S). For the decoding analysis we used time lags between S and R ranging from −400 to 0 msec (negative lags imply that the S precedes the R; here including a baseline would be detrimental to decoding). The decoding model was optimized using the same method and ridge parameter as the encoding model, using a regularization parameter of $\lambda = 100$ and a leave-one-out procedure to test the predictive power of the model. For the decoding approach, the predictive power is the Pearson's correlation between the reconstructed S using the model and the actual S. Since the decoding model results in a single value for each participant, it is somewhat simpler to use for statistical comparison across groups and conditions, as no corrections for multiple comparisons are needed. Between-group differences in speech-decoding accuracies were assessed using unpaired t-tests.

Event Related-Response

To quantify neural responses to the sound-events, we estimated the time-locked Event-Related Potentials (ERPs). For this analysis, the clean EEG data was segmented into epochs ranging from -100 to 500ms around each event. The data was further low-passed filtered at 12Hz, as is common in ERP analysis and baseline corrected to the pre-stimulus period (-100 to 0 ms). ERPs were derived separately for each group (ADHD vs. Control) and Event-type (Artificial vs. Non-verbal human sounds). Visual inspection of the ERPs showed that they were dominated by two early components: the N1-component (75-200ms) and later P2-component (210-260ms), which primarily reflect auditory sensory responses, but are sometimes modulated by attention (Hillyard et al., 1973; Woldorff and Hillyard, 1991).

Statistical analysis of the ERPs focused on the two simple effects, using a Monte-Carlo spatio-temporal cluster-based permutation test, across all electrodes and time points. To test for difference in the ERPs between Groups, we used unpaired t-tests on the ERPs, averaged across Event-types). For this analysis, we had to set the cluster alpha-level to a very low level of $p=10^{-5}$, in order to obtain separate clusters for the N1 and P2 responses. To test for difference in the ERPs generated for the two Event-types, we use paired t-tests on the ERPs of all participant, irrespective of their group. For this analysis, the cluster alpha-level was set to p=0.05. Last, to verify the main effects and in order to test for potential interactions between Group and Event-type, we extracted the peak-amplitudes of the N1 and P2 component from each participant, and performed a Mixed ANOVA with repeated measures, with the factors Group (ADHD vs. Control; between) and Event-Type (Artificial vs. Human non-verbal; within).

Spectral Analysis

The third analysis performed on the EEG data was a spectral analysis. This analysis was performed on the clean EEG data, segmented into full trials (same segmentation used for the speech-tracking analysis). We calculated the EEG spectral power density (PSD) of individual trials using multitaper fast-fourier transform (FFT) with Hanning tapers (method 'mtmfft' in the fieldtrip toolbox). The PSDs were averaged across trials for each participant, separately for each electrode and condition (Quiet and Events). We used the FOOOF algorithm (Donoghue et al., 2020) to decompose the PSD into periodic (oscillatory) and aperiodic components. The periodic portion of the PSD showed clear peaks in the Alpha (8-12 Hz) and Beta (15–25 Hz) ranges, suggesting

reliable oscillatory activity (Figure 7). Since the peaks of Alpha and Beta activity can vary substantially across individuals, for each participant, we identified the frequency with the largest amplitude within each range, focusing on a cluster of occipital-parietal electrodes (for Alpha) or frontal-central electrodes (for Beta), where these responses were largest. These were used for statistical analyses of between-group differences. assessed using unpaired t-tests, separately for the Quiet and Events conditions, in each frequency band, and the Bayes Factor (BF_{10}) was calculated for each pairwise comparison.

Skin Conductance data analysis

The Skin Conductance (SC) signal was analyzed using the Ledalab MATLAB toolbox (Benedek and Kaernbach, 2010) and custom-written scripts. The raw data was downsampled to 16Hz and inspected for any noticeable artifacts, which were corrected using linear interpolation. A continuous decomposition analysis (CDA) was employed to separate the tonic and phasic SC responses for each participant. We tested for between-group differences in *global SC levels* (tonic and phasic responses separately, averaged across full trials) using unpaired t-tests, separately for the Quiet and Events conditions and the Bayes Factor (BF_{10}) was calculated for each pairwise comparison.

In addition, we analyzed *event-related changes in SC* following Sound-events. This analysis focused on a time window between 0 to 5 seconds after the onset of Sound-events. Since changes in SC are relatively slow and peak between 2-3 seconds after a stimulus, the mean response between 0-1 seconds was used as a baseline and was subtracted from the signal. Statistical analysis focused the time-window surrounding the mean response between 2-3 seconds, extracted for each participant. A Mixed ANOVA with repeated measures was used to test for main effects of Group (ADHD vs. Control; between factor), and Event-Type (Artificial vs. non-verbal Human; within factor) and the interaction between them.

Eye tracking data analysis

Eye-tracking data was used to study participants' spontaneous gaze dynamics in the VR classroom throughout the entire trial and also to examine whether background Sound-events elicited overt gaze shifts. Gaze-data preprocessing involved removing measurements around blinks (from -100ms to 200ms after) and any other data points marked as "missing data" by the eye-tracker. Gaze-data were analyzed using in-house written Python scripts, and closely followed

recommended procedures for analysis of gaze-data in VR (Anderson et al., 2023). To assess gaze-patterns and detect gaze-shift, we calculated the Euclidean distance between adjacent data points, using the x, y, z coordinates in 3D space provided by the eye-tracker. 'Fixations' were defined as clusters of adjacent time-points within a radius of less than 0.01 distance from each other, lasting at least 80 ms. Each fixation was associated with a specific ROI in the VR classroom, indicating the object in the virtual space that the participant looked at. 'Gaze-shifts' were defined as transitions between two stable fixations.

To analyze spontaneous gaze-dynamics throughout an entire trial, we quantified the percent of each trial that participants spent looking at the teacher (target) versus all other ROIs, as well as the number of gaze-shifts away from the teacher per trial. Between-group differences in these measures of spontaneous gaze patterns were assessed using unpaired t-tests, separately for the Quiet and Events conditions.

To determine whether background Sound-events elicited overt gaze shifts, we segmented the gaze-data into 2-second long epochs around each Event-sound. As a control, we randomly chose a similar number of 2-second long epochs from the Quiet-condition. We then counted how many epochs included at least one gaze-shift away from the teacher, and used a paired t-test to determine whether the frequency of gaze-shifts was higher following Sound-events relative to the Control epochs (regardless of Group). We also used a Mixed ANOVA to test for potential differences in the likelihood of gaze-shifts following the different event types (Artificial vs. Human non-verbal), and for the two Groups (ADHD vs. Control). For this analysis, the proportion of gaze shifts was normalized for each participant by subtracting the number of gaze shifts following Sound-events relative to the control epochs.

Exploratory Correlational Analyses

Besides testing for effects of Group in each metric individually, we performed two exploratory analyses to test the relationship between the plethora of neurophysiological metrics collected here, and whether they predict the severity of ADHD symptoms, as reflected by ASRS scores (self-report). Exploratory analyses included evaluating the relationship between all metrics (Spearman correlation) and a multiple linear regression analysis testing which metrics carried substantial explanatory power of the variability in ASRS scores across individuals (dependent measure). We used a backwards hierarchical approach which helps distill which factors had the best explanatory

power of ASRS score variability and reduces the analysis to the model with the best explanatory power and the least number of factors. Both analyses were carried out using the software JASP.

Results

Behavioral results

Participants demonstrated overall good performance on the comprehension task, achieving an average accuracy of 87.87% (±6.77% SEM). This serves as verification that participants followed the instructions, listened to teacher's speech and understood the content of the mini-lessons. Performance levels were similarly good in both the Events and Quiet conditions, indicating that the presence of occasional sound-events did not disrupt overall understanding of the mini-lessons. No significant differences in performance were found between groups, in either condition [Events: $t(47) = 1.052$, $p = 0.297$, Cohen's $d = 0.30$, $BF_{10} = 0.448$ (weak evidence for H0); Quiet: $t(47) = 1.004$, $p = 0.320$, Cohen's $d = 0.28$, $BF_{10} = 0.430$ (weak evidence for H0); Figure 3].

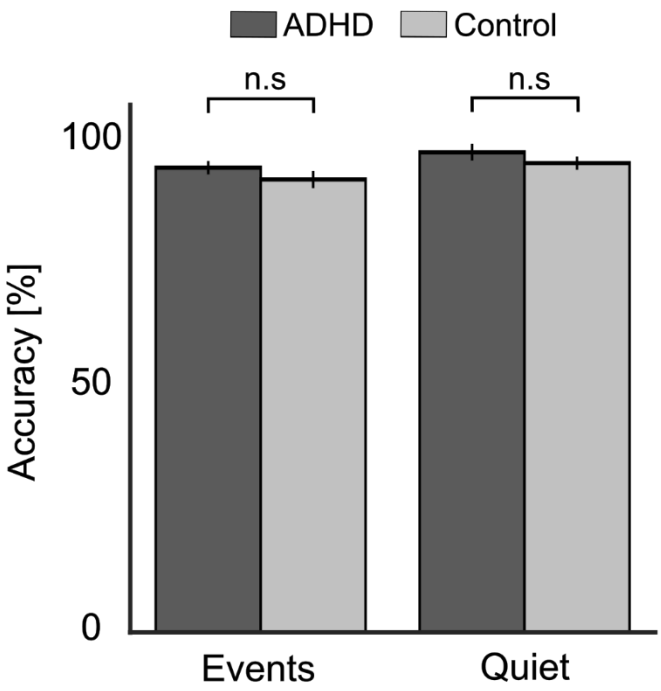

Figure 3. Accuracy on answering the multiple-choice comprehension questions, shown separately for the ADHD and Control groups, in the Quiet and Events conditions. Bar graphs represent average accuracy levels across participants and error bars represent the SEM. No significant difference was found between the ADHD and Control groups (n.s).

EEG results

Speech tracking response

We conducted speech-tracking analysis of the neural response to the teacher's speech, using an encoding approach. Before comparing between the groups, we assessed the suitability of using a univariate vs. multivariate encoding model to quantify the response to speech in the Events condition, that also included sound-events. Figure 4A shows that qualitatively similar TRFs were estimated using both approaches, with a negative peak around 100ms followed by a positive peak at 150ms, similar to the pattern observed in previous studies (Lalor and Foxe, 2010; Zion Golumbic et al., 2013; Ding and Simon, 2014; Kaufman and Zion-Golumbic, 2023). The TRFs estimated in the Events condition were qualitatively similar to the TRF estimated in the Quiet condition (based only on n=8 trials), which did not contain additional event-sound, lending additional credibility to the approach (eventhough given the large difference in number of trials between conditions, TRFs in the Quiet and Events conditions could not be reliably compared statistically). In the multivariate model, a TRF capturing the response to sound-events was also estimated (Figure 4B), which also shows two prominent peaks - negative at 100ms and positive at 200 - a pattern comparable to the ERPs derived for these stimuli through simple averaging (see Figure 6 below), as expected for transient stimuli (given the linear nature of the model. When comparing the predictive power of the univariate and multivariate models, both models yielded significant predictive power relative to null-permutations, however the multivariate model was able to explain a larger degree of the variance in the neural response (Figure 4C). This is probably because including a second regressor with the audio of the sound-events allows the model to more faithfully capture the neural representation of the entire VR classroom soundscape (Crosse et al., 2021). This comparison carries important methodological implications for future studies of attention in realistically noisy environments, and given these results, further statistical analysis focused on the results from the multivariate model.

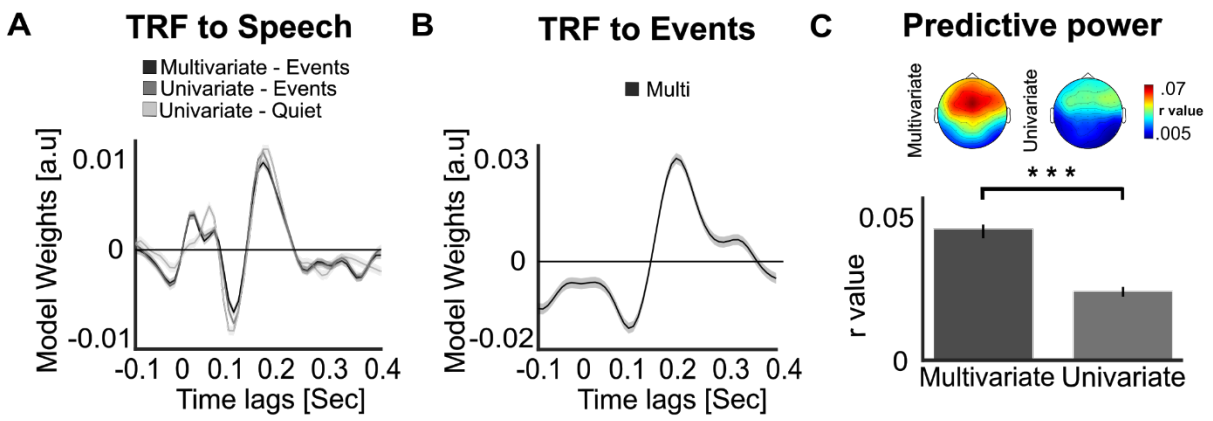

Figure 4. Comparison between Univariate and Multivariate TRF model. A. Grand-average TRFs (across both groups) reflecting the neural tracking response of the teacher's speech, estimated for the Events condition using both a univariate and multivariate model, as well as for the Quiet condition (univariate model). Shading represent the standard error of the mean (SEM) **B.** Grand-average TRF estimated for the regressor representing the audio of the event-sounds, in the multivariate model of the Events condition. Shaded areas represent the SEM. This analysis is akin to calculating the event-related potential (ERP) to event-sounds, shown in Figure 6. **C.** Topographical distribution (top) and bar graph (bottom) depicting the predictive power values (Pearson's r) in the Events condition for the univariate and multivariate encoding models.

We then turned to comparing the neural speech tracking response between the two groups, focusing on the Events condition only. The encoding analysis did not reveal any significant differences in predictive power between the groups (Figure 5A), nor were any significant spatio-temporal clusters identified when comparing the TRFs estimated in each group (Figure 5B). Interestingly, the decoding analysis did show a slight decrease in the reconstruction accuracy of the stimulus in the ADHD group relative to controls [$t(47) = -2.023$, $p = 0.048$, Cohen's $d = -0.578$, $BF_{10} = 1.47$ (weak evidence for H1); Figure 5C].

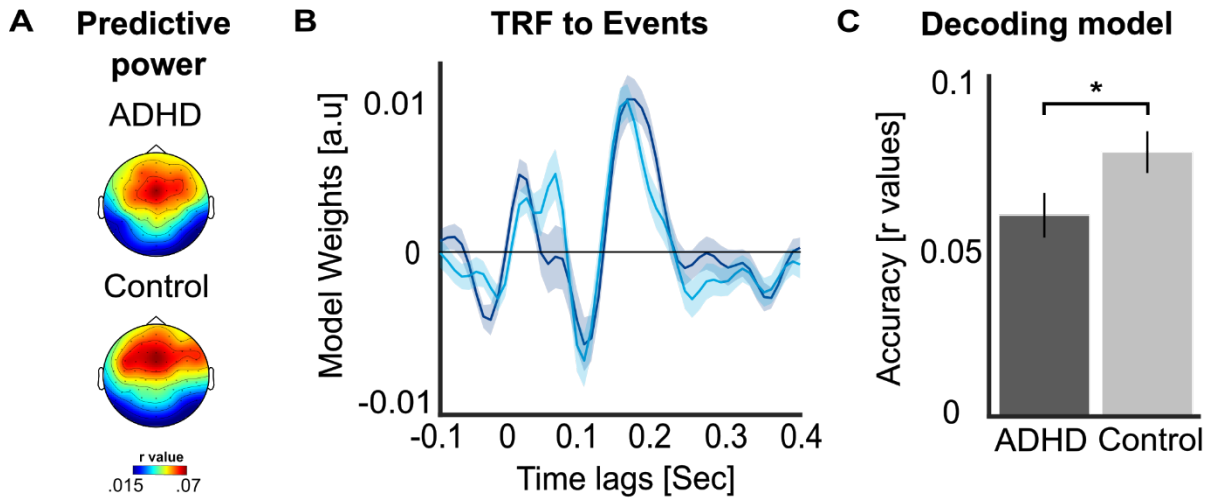

Figure 5. Between-group comparison of the neural speech tracking response in the Events-conditions. A. Topographical distribution of the predictive power values, estimated using the multivariate encoding model, in the ADHD and Control groups**.** No significant differences were observed between groups. **B.** Grand-average TRFs estimated for each Group in the Events condition. Shaded areas represent the SEM. **C.** Bar graphs showing the average reconstruction accuracy (Pearson's r) of the decoding model, in each Group. Asterisk (*) indicate a significant difference between ADHD and Control groups ($p < 0.05$). Error bars represent the SEM.

Event Related Potentials (ERPs)

ERPs were derived response to sound-events, and we tested for potential differences between the two Groups (ADHD vs. Control), and the two Event types (Artificial vs. non-verbal Human sounds). Visual inspection of the ERPs showed two prominent centro-parietal components - a negative peak around 100ms, which corresponds to the early sensory N1; followed by a positive peak around 240ms, which corresponds to the P2 (Figure 6).

Statistical analysis comparing ERPs in the ADHD vs. Control groups (collapsed across both Event-Types) showed that the early N1 response was significantly larger in the ADHD group ($p<0.05$, cluster correction; Figure 6A&B). No significant differences between groups were found for the later P2 response. When comparing ERPs to Artificial vs. non-verbal Human sounds we found that both the N1 and the P2 were significantly modulated by Event-type, with a larger N1 for the Artificial sounds and a larger P2 for the non-verbal Human sounds (both $p<0.002$, cluster correction; Figure 6C&D). A mixed-ANOVA performed on the amplitudes of each component confirmed these main effects, but did not reveal any interaction between Group and Event-type

[N1: *Group* [F(1, 47) = 6.15, p = 0.044, η^2 = 0.11], *Event type* [F(1, 47) = 63.82, p < 10^{-9}, η^2 = 0.57], *interaction* [F(1, 47) = 0.389, p = 0.24, η^2 = 0.008]; P2: *Group* [F(1, 47) = 0.84, p = 0.36, η^2 = 0.017], *Event type* [F(1, 47) = 71.75, p < 10^{-9}, η^2 =0.6], *interaction* [F(1, 47) = 0.07, p=0.78, η^2 = 0.001; Figure 6].

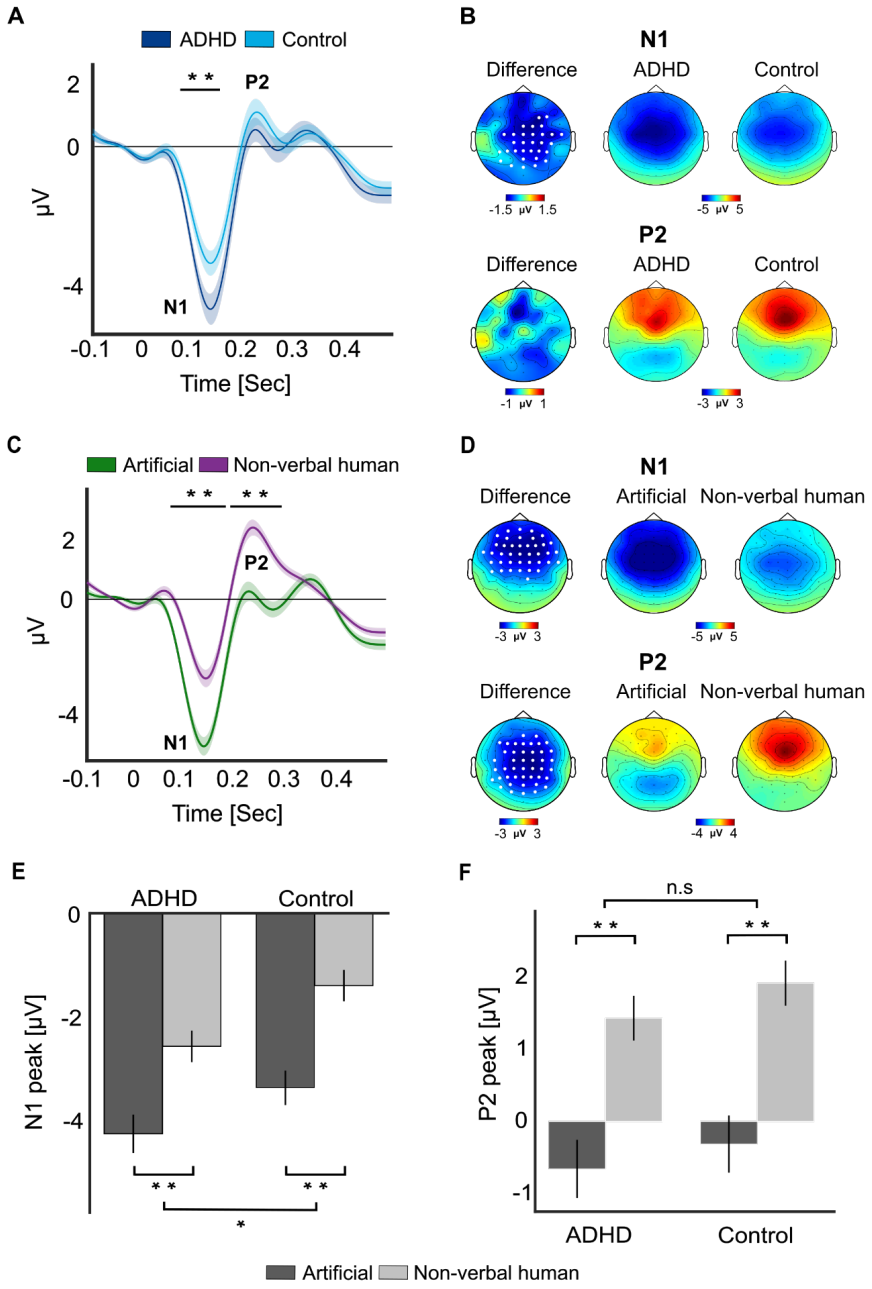

Figure 6. Event Related Response to Sound-events. A. Grand-average ERP to sound-events, shown separately for the ADHD vs. Control groups. ERPs are averaged across the cluster of electrodes where significant differences were found (see panel B). The black horizontal line indicates the time-window where significant differences between groups were found (75-155ms). Shaded areas around the waveforms represent the SEM. **B.** Scalp topographies of the N1 and P2 responses in the ADHD and Control Groups, and the difference between them. Electrodes where significant differences between groups were found are marked in white ($p<0.05$, cluster corrected). **C.** Grand-average ERP to Artificial and non-verbal human event-sound. ERPs are averaged across the cluster of electrodes where significant differences were found (see panel D). The black horizontal lines indicates the time-windows where significant differences between ERP to the two types of sound-events were found (67-178 and 187-291ms). Shaded areas around waveforms represent the SEM. **D.** Scalp topographies of the N1 and P2 responses to Artificial and Non-verbal human sounds and the difference between them. Electrode where significant differences were found are marked in white ($p<0.05$, cluster corrected). **E&F.** Box plots depicting the average N1 and P2 responses, separately for each group (ADHD vs. Control) and Event-type (Artificial vs. Non-verbal human). Box plots represent median values, error bars represent the interquartile range (1.5 x IQR) and small "x's" represent outliers outside this range. ** $p < 0.001$; * $p<0.05$.

Spectral EEG Analysis

Spectral analysis of the EEG focused on the two frequency bands for which clear peaks were detected in the periodic power-spectrum, after applying the FOOOF algorithm: Alpha (8-12Hz) and Beta power (18-25Hz; Figure 7A). Power in each frequency band was assessed for each participant at their personal peak within a pre-selected cluster of electrodes where each response was maximal (Figure 7B). We then tested whether power in either frequency differed significantly between Groups, using unpaired t-tests, separately for each Condition (Quiet and Event trials). However, none of the comparisons revealed any significant differences between groups in the power of ongoing oscillatory responses (Figure 7C&D) [*Alpha Power*: Events Condition: $t(47) = -0.656$, $p = 0.514$, Cohen's $d = -0.18$,$BF_{10} = 0.34$ (weak evidence for H0); Quiet Condition: $t(47) = -0.394$, $p = 0.695$, Cohen's $d = -0.11$, $BF_{10} = 0.30$ (moderate evidence for H0); *Beta-power:* Events Condition: $t(47) = -1.01$, $p = 0.315$, Cohen's $d = -0.29$, $BF_{10} = 0.43$ (weak evidence for H0); Quiet Condition: $t(47) = -0.484$, $p = 0.630$, Cohen's $d = -0.14$, $BF_{10} = 0.314$ (moderate evidence for H0)] (Figure 7D).

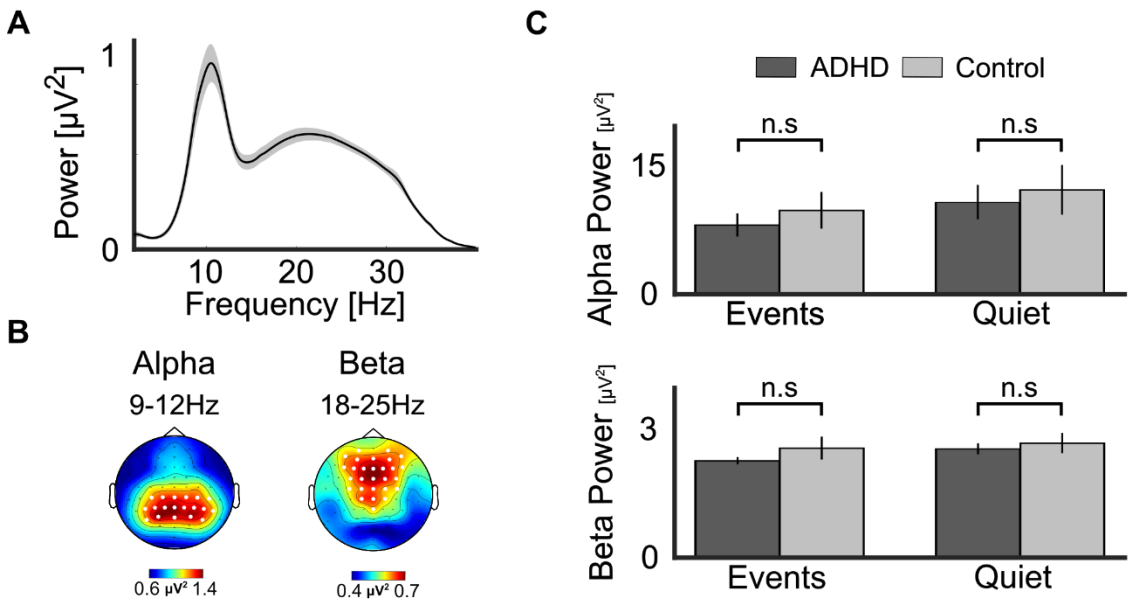

Figure 7. Spectral EEG Analysis. A. Grand-Average Power spectral density (PSD) of the Periodic portion of the EEG signal (after applying the FOOOF algorithm), shows two clear peaks corresponding to the alpha (8-12Hz) and beta (15-25Hz) bands. Shaded areas around waveforms represent the SEM. **B.** Topographical distribution of the average alpha-power and beta-power peaks, with the clusters of electrodes used to detect personal peaks in each frequency band marked by white circles. **C&D.** Average alpha- and beta-power in Group (ADHD vs. Control) and condition (Quiet and Events). Bar graphs represent the mean values, and error bars represent the SEM. No significant between-group difference was found in any comparison (n.s.).

Skin Conductance (SC) results

Global: No significant differences between the two Groups were observed for the global tonic or phasic SC metrics, in either the Events or Quiet condition [***Phasic SC*** (Figure 8A): Events Condition*:* $t(47) = -0.003$, $p = 0.99$, Cohen's $d = -0.0009$, $BF_{10} = 0.285$ (moderate evidence for H0); Quiet Condition: $t(47) = -0.51$, $p = 0.61$, Cohen's $d = -0.146$, $BF_{10} = 0.317$ (moderate evidence for H0) ; ***Tonic SC***: Events Condition*:* $t(47) = -0.85$, $p = 0.398$, Cohen's $d = -0.244$, $BF_{10} = 0.383$ (weak evidence for H0); Quiet Condition: $t(47) = -1.65$, $p = 0.104$, Cohen's $d = -0.476$, $BF_{10} = 0.865$ (weak evidence for H0)].

Event-Related: Sound-events elicited clear event-related changes in the phasic SC response, which peaked 2-3 seconds after the sounds, and took another ~1 second to return to baseline (Figure 8B & C). A Mixed ANOVA (Figure 8D) showed a significant main effects of Event type, with a higher response following artificial sounds relative to non-verbal human sounds

[F(1, 47) = 7.88, p = 0.007, η^2 = 0.143], but no significant main effect of Group [F(1, 47) = 0.4, p = 0.53, η^2 = 0.008] or interaction between them [F(1, 47) = 0.02, p = 0.87, η^2 = 0.0005].

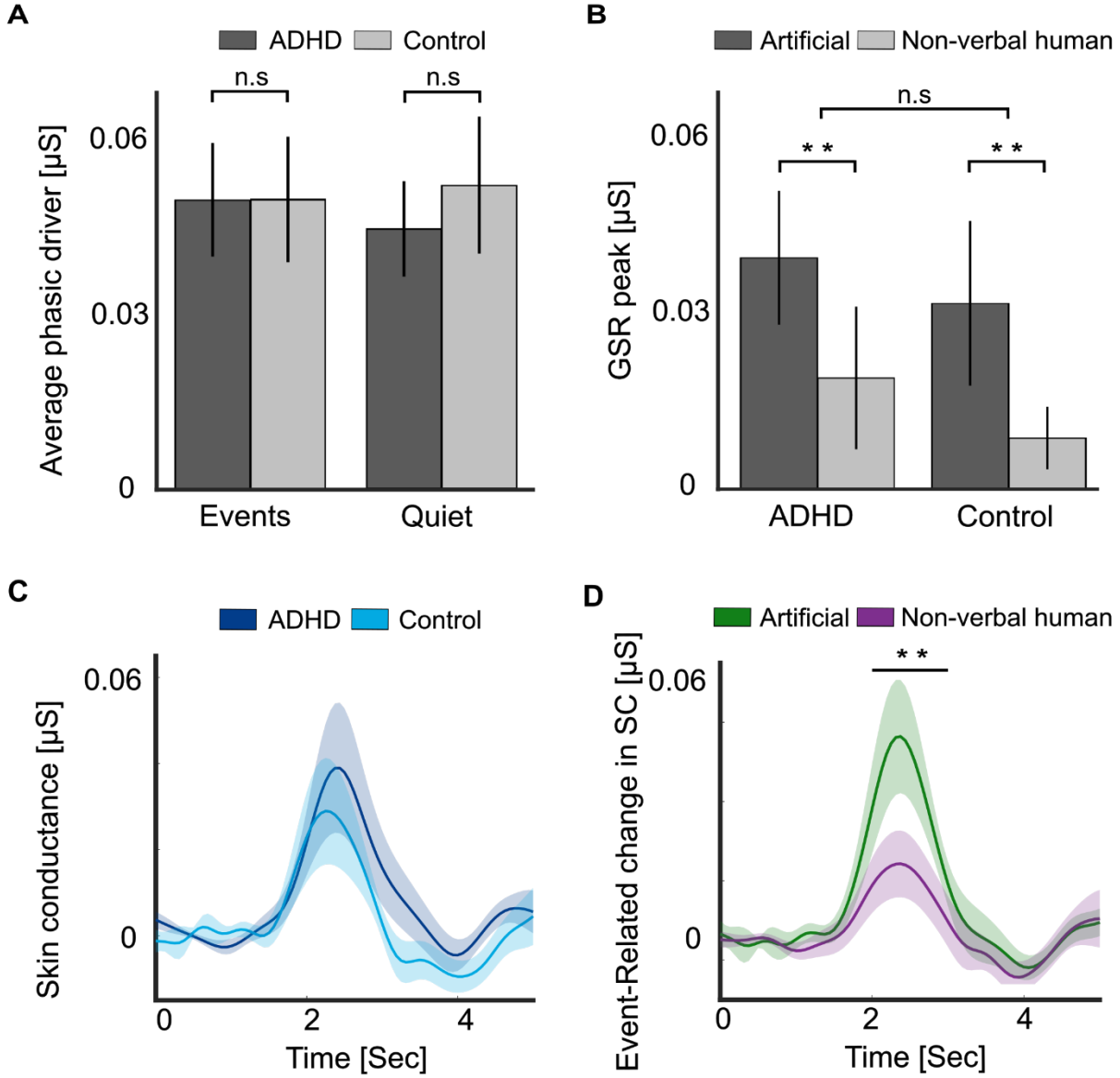

Figure 8. Skin Conductance results. A. Average levels of phasic SC levels in the ADHD and Control groups, shown separately in the Events and Quiet conditions. The error bars represent SEM. No significant between group differences were found in any comparison (n.s.). **B-D** Event-related changes in SC. **B.** Average levels of event-related changes in phasic SC following event-sounds (peak between 2-3sec after the event; see panels C&D), shown separately for each group and for Artificial vs. Non-verbal human sound-events. Error bars represent the SEM. **p<0.01. **C.** Time course of event-related changes in phasic SC following the sound-events, shown separately for each group. Shaded areas around waveforms represent the SEM. No significant between group differences were found in e. **D.** Time course of event-related changes in phasic SC following the two types of sound-events (Artificial vs. Non-verbal human) averaged across groups. Shaded areas around waveforms represent the SEM. Horizontal line represents the time-windows where significant differences between event types were found.

Eye Gaze results

Global: Analysis of global gaze-patterns throughout an entire trial revealed that, as expected, participants spent most of the time looking at the teacher (60.2% ± 20.9% SEM of each trial). However, we also note the large variability across participants, with some focusing almost exclusively on the teacher (near 100%) and others spending less than 40% of the trial looking at the teacher. When not looking at the teacher, the next most popular places to look were the blackboards behind the teacher, to the right and to the left (Figure 9A). We tested whether the amount of time spent looking at the teacher and the number of gaze-shifts away from the teacher were modulated by Group (ADHD vs. Control) separately in each Condition (Quiet and Events). However, no significant differences were found [***Percent gaze-time at teacher:*** Events Condition: $t(47) = -0.899$, $p = 0.372$, Cohen's $d = -0.257$, $BF_{10} = 0.397$ (weak evidence for H0); Quiet Condition: $t(47) = -0.597$, $p = 0.553$, Cohen's $d = -0.170$, $BF_{10} = 0.33$ (moderate evidence for H0) ; ***Number of gaze-shifts away from teach*:** Events Condition*:* $t(47) = 1.265$, $p = 0.211$, Cohen's $d = 0.361$, $BF_{10} = 0.547$ (weak evidence for H0); Quiet Condition: $t(47) = 0.644$, $p = 0.522$, Cohen's $d = 0.184$, $BF_{10} = 0.338$ (weak evidence for H0) ; Figure 9C].

Event Related Gaze-Shifts: We tested whether the sound-events triggered overt gaze-shifts away from the teacher, by quantifying the percentage of event-sound that were followed by at least one gaze-shift, and comparing this to control epochs taken from the Quiet condition. However, no significant differences were found, suggesting that the sound-events did not elicit more frequent overt gaze-shifts away from the teacher, relative to what might be observed when no sound-events are present [$t(48) = 1.34$, $p = 0.18$]. We also tested whether the likelihood of performing event-related gaze-shifts was different for the two Event types (Artificial vs. non-verbal human sounds) or in the two Groups (ADHD vs. Control) but did not find any significant main effects or interactions [*Group*: $F(47) = 0.18$, $p = 0.67$, $\eta^2 = 0.003$; *Condition*: $F(47) = 2.25$, $p = 0.14$, $\eta^2 = 0.045$; *Interaction*: $F(47) = 0.51$, $p = 0.47$, $\eta^2 = 0.01$; Figure 9D].

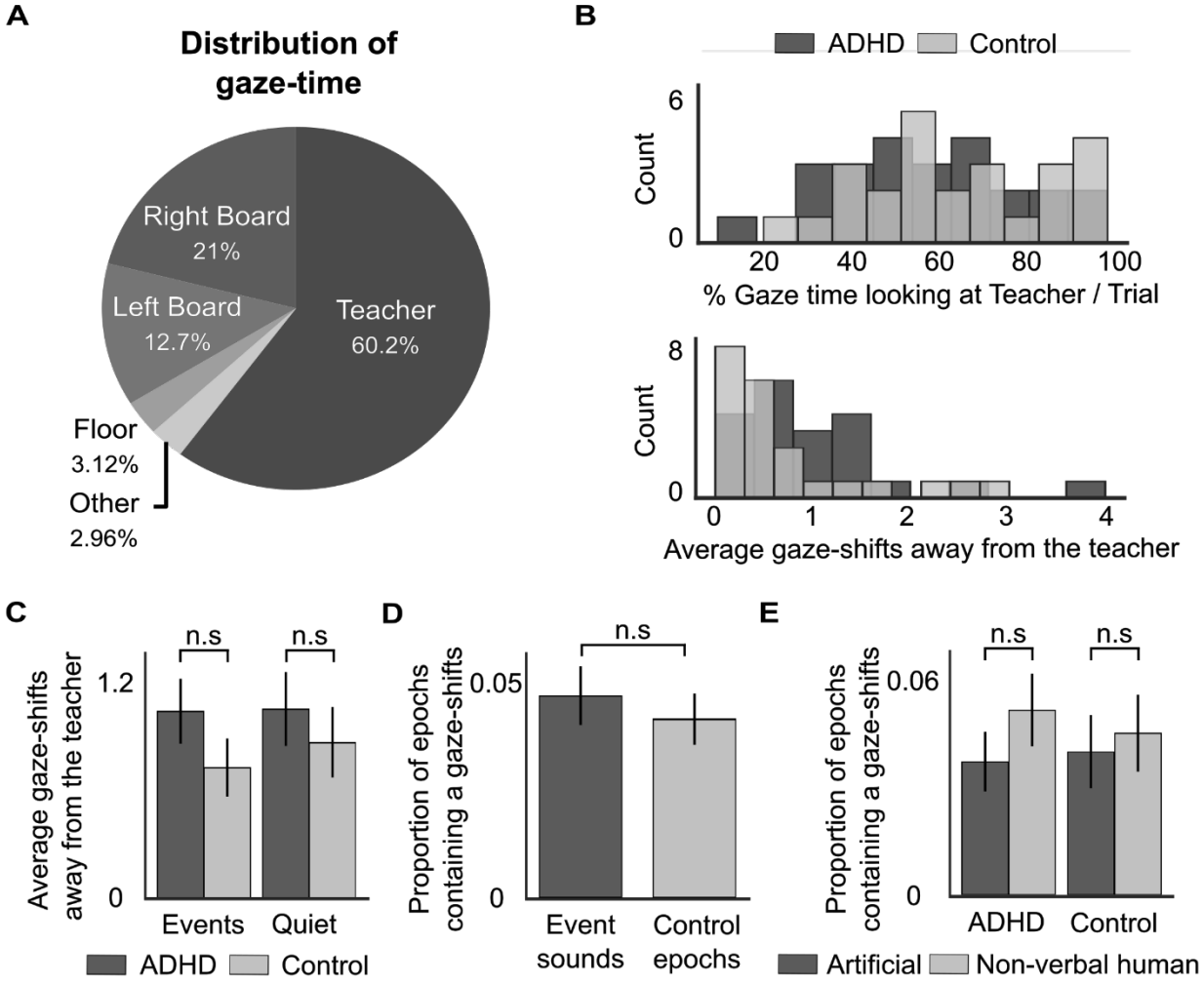

Figure 9. Eye-gaze results. A. Pie chart representing the average amount of time that participants spent looking at different areas within the VR classroom. **B.** Distribution of the proportion of gaze-time towards the teacher (top) and number of gaze-shifts performed away from the teacher (bottom), for all participants in the ADHD and Control groups. **C.** Bar graph represent the average number of gaze-shifts performed away from the teacher, shown separately for the ADHD and Control groups, and for the Quiet and Events conditions. No significant differences were found in any comparison (n.s) **D.** Bar graph showing the number of gaze-shifts performed in 2-sec epochs following event-sounds vs. control epochs, averages across groups and sound types. No significant differences were found, suggesting that event-sounds were not more likely to elicit overt gaze-shifts. **E.** Bar graph showing the number of gaze-shifts performed in 2-sec epochs following each type of sound-event, separately for each group. No significant differences were found in any comparison. Error bars represent the SEM in all bar graphs.

Exploratory Correlational Analyses

Besides testing for effects of Group in each metric alone, we performed exploratory analyses testing the relationship between the plethora of neurophysiological metrics collected here, and their ability to predict the severity of ADHD symptoms, as reflected by ASRS scores (self-report). These analyses focused initially on 10 independent metrics (averaged across conditions, when relevant): Behavioral performance, neural speech tracking decoding accuracy, amplitude of the N1-response and P2-response to sound-events, alpha-power, beta-power, tonic SC response, phasic SC response, the overall time spent looking at the teacher and the number of spontaneous gaze-shifts away from the teacher.

We first evaluated the Spearman correlation between all metrics and found that the two SC-based metrics (phasic and tonic responses) and the two gaze-based metrics (percent gaze-time towards teacher and number of spontaneous gaze-shifts) were highly correlated with each other (Spearman's rho = 0.63 and rho = -0.78, respectively, $p<0.001$ uncorrected, $p<0.02$ fdr corrected; for both; Figure 10A). Therefore, for our multiple regression model, we chose to include only one of each – phasic SC response and number of spontaneous gaze-shifts away from the teacher. Two other metrics were modestly correlated with each other: N1 amplitude and phasic SC (Spearman's rho = 0.3, $p=0.036$ uncorrected); P2 amplitude and neural speech decoding accuracy (Spearman's rho = -0.33, $p=0.02$ uncorrected). However, these results did not survive fdr correction for multiple comparisons, and therefore require replication in subsequent studies.

We next performed a multiple linear regression analysis, with ASRS score as the dependent factor and the remaining 8 measures as independent covariates. We used a backwards hierarchical approach which helps distill which factors had the best explanatory power of ASRS score variability and reduces the analysis to the model with the best explanatory power and the least number of factors. The final regression model identified three factors that best explained ASRS scores, with an overall explanatory power of 19% of their variance [$r = 0.43$, $r^2 = 0.19$; $F(3,34)=3.54$, $p=0.02$]. This model included N1 amplitude [$\beta = 0.28$, $t(45)=2.08$, $p<0.05$]; number of spontaneous gaze-shifts away from the teacher [$\beta = 0.29$, $t(45)=2.12$, $p<0.05$] and alpha power [$\beta = 0.28$, $t(45)=2.03$, $p<0.05$]. All three factors were positively correlated with ASRS scores (Figure 10B).

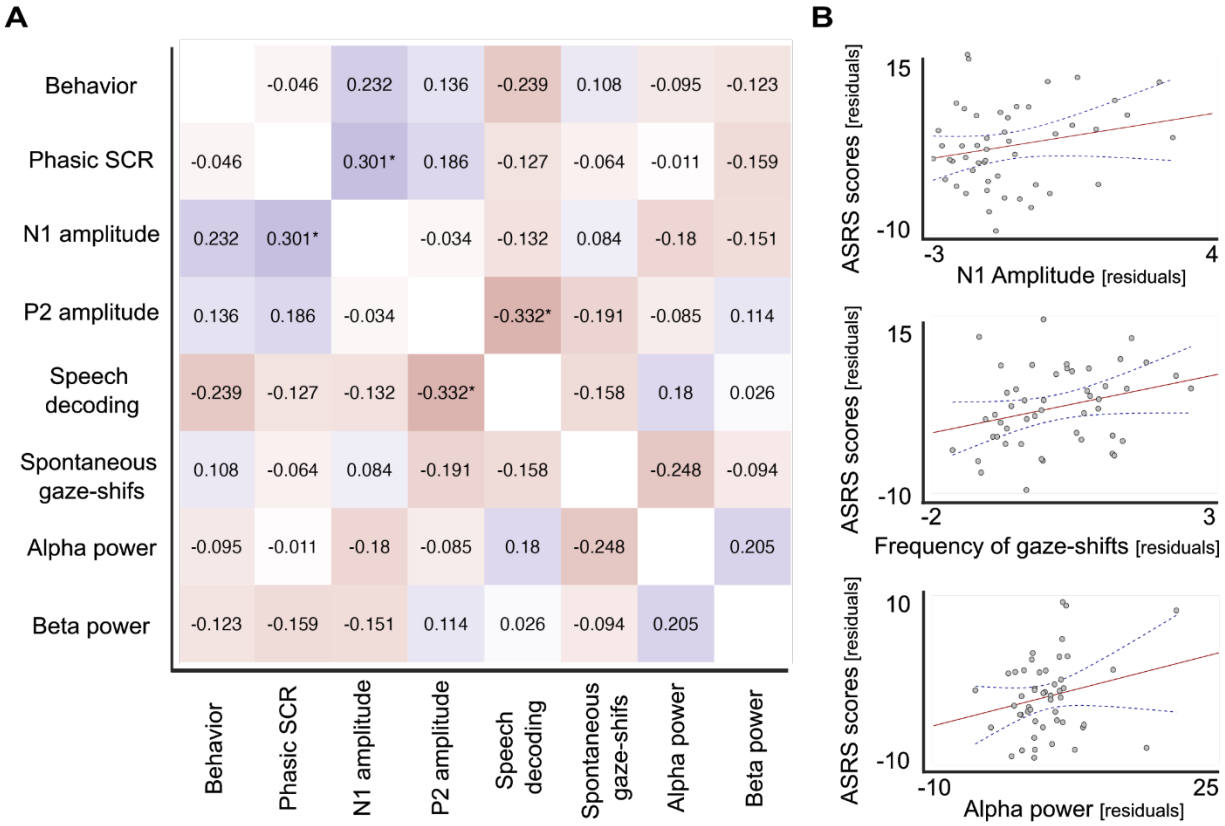

Figure 10. Exploratory correlational analysis A. Heat-map showing the Spearman's correlation values (r) between every pair of metrics, and color coded to reflect the direction (red-negative; blue-positive) and strength of the correlation (color intensity). Correlations marked with an asterisk were significant ($p<0.05$ uncorrected), but none survived fdr correction for multiple comparisons. **B.** Scatter plots showing the correlation of between N1 amplitude (top), frequency of spontaneous gaze shifts (middle) and alpha power (bottom) with ASRS scores. Values are shown as residuals from the multivariate regression model. The regression line is shown in red and dashed lines surrounding it represent confidence intervals. These three factors were identified in a backwards hierarchical multivariate regression as having the best explanatory power of ASRS scores.

Discussion

The rich multivariate dataset collected here breaks new ground in providing a platform for studying the neurophysiological mechanisms of attention and distraction under highly ecological conditions. We focused on VR classroom learning, since this is the context where many individuals, and particularly those with ADHD, often experience difficulties in sustaining attention and avoiding distraction. Indeed, we show that irrelevant unexpected background sound-events elicited neural ERPs and a transient increase in SC, in both groups, indicating that these

disturbances impact neural processing and arousal levels (although they did not capture overt attention). Moreover, the early N1 neural responses to these sounds was significantly larger in the ADHD group, suggesting they exhibit increased sensory sensitivity to irrelevant background stimuli. In addition, between-subject variance of the N1 amplitude, power of alpha oscillations and the frequency of spontaneous gaze-shifts around the classroom had significant explanatory power of the variance in self-reported ADHD symptoms across the sample. These three neurophysiological measures have been associated with different aspects of inattention/distractibility in more traditional laboratory-based research, and the current results provide ecological support for their more prevalent presentation in ADHD.

At the same time, many metrics did not differ significantly between groups. Most notable perhaps is that both groups performed the task well, and showed similar neural tracking of the teacher's speech. This raises questions about how 'distracting' the sound-events really were and suggests that noticing/responding to irrelevant stimuli might not always detract from performing the main task. Moreover, counter to prior suggestions, we do not find evidence for differences in levels of arousal and altered beta-band neural oscillations between the groups, challenging the notion that these may constitute neurophysiological 'markers' of ADHD. These data resonate with recent calls to revisit the parameters used for defining 'attention-disorders' and the need to adopt more nuanced and dimensional approach for characterizing the nature of real-life attentional challenges (Marcus and Barry, 2011; Heidbreder, 2015; Morris et al., 2022; Elahi et al., 2023).

In discussing these results and their implications we bear in mind that – like any experiment – they are limited to the specific features and stimulus-choices of the current VR classroom. Moreover, given the many neurophysiological metrics analyzed here, and the exploratory nature of some analysis, the strength of this work lays primarily in the new hypotheses it brings to the table regarding attention and distraction during ecological learning and the neurophysiological metrics that are relevant for understanding these constructs.

Event-related Response to Sound Events

Sound-events in the background of the VR classroom elicited two well-known neurophysiological responses: auditory ERPs and transient changes in SC. ERPs showed the typical P1-N1-P2 ERP complex, that reflects a cascade of neural responses to sounds in primary and association auditory cortex, peaking approximately at 50, 100 and 150ms respectively

(Hillyard et al., 1973; Näätänen and Picton, 1987; Woldorff and Hillyard, 1991; Woldorff et al., 1993). SC responses are consistent with the well-known 'orienting reflex', which peaks ~2-3 seconds after an event, and is associated with a transient event-induced increase in arousal (Frith and Allen, 1983; Zimmer and Richter, 2023). Both responses are considered obligatory and automatic, as they do not require allocation of attention and can even be observed during sleep and other unconscious states (albeit the magnitude of these responses can vary as a function of attention). Previous studies have provided conflicting evidence regarding the connection between neural responses to sounds and the autonomic orienting-reflex (Lim et al., 1996; MacDonald and Barry, 2020), however in the current data these two metrics were positively correlated. Together this pattern indicates that all sounds in the environment generate automatic sensory responses, even if they are task-irrelevant, and lead to momentary increases is arousal – which may be the basis for distraction.

We further find that the N1 response to sound-events was larger in the ADHD group relative to controls and that variability in N1 amplitude contributed to explaining the variability in ADHD symptoms across participants. This was somewhat surprising, since most studies comparing auditory ERPs in individuals with and without ADHD, using more traditional paradigms (e.g. oddball), typically *have not reported* differences in early-sensory components such as the N1. If anything, between-group differences are sometimes reported for later ERP responses, such as the P300 and the Late Positive Component (LPC), which have been associated with capture of attention and distractibility by surprising stimuli, but not to early sensory processing (Barry et al., 2003; Gumenyuk et al., 2005; Kaiser et al., 2020; Peisch et al., 2021). In the current study these later components were not clearly visible in the ERP, raising important questions as to their generalizability to more ecological contexts (see further discussion below). That said, enhanced N1 responses are often associated with heightened sensory sensitivity (Heil, 1997; Lütkenhöner and Klein, 2007), perceptual predictability (Chennu et al., 2013; Lange, 2013; Schwartze and Kotz, 2015), and may reflect a transient detector mechanism of surprising events that can potentially (though not necessarily) trigger an attention shift (Näätänen and Picton, 1987; Escera et al., 1998, 2003; Berti, 2013). The current finding may suggest that individuals with ADHD, and particularly those who report more severe symptoms, exhibit higher sensory sensitivity to background sound-events. However, this effect alone does not necessarily reflect distraction by these stimuli.

When comparing the ERP and SC responses for the two types of sound-events, we found similar time-course for both measures, but a different magnitude of response to the artificial sounds relative to the non-verbal human sounds (coughs). These differences may be due to acoustic differences between the sounds, since although all stimuli were equated for overall loudness and duration, the artificial sounds had sharper onsets and offsets relative to the cough-stimuli which were more continuous. Alternatively, the difference in response may be due to the semantic/contextual differences between them, since the non-verbal human sounds are more contextually relevant to a classroom than artificial 'boing' sounds. Since the current study was not designed to systematically investigate the effects of specific acoustics and semantic parameters, additional follow-up research is needed to better understand the relative contribution of these factors to the neurophysiological responses to background sounds. However, the current results do indicate that the system is highly sensitive no only to the presence of background sounds but to features other than simply their intensity. They also demonstrate that as attention research incorporates increasingly ecological stimuli, untangling the effects of acoustic vs. semantic features becomes increasingly difficult, given the nature of the stimuli.

Detection or Distraction?

Does the finding that sound-event elicited neural and SC responses indicate that individuals with ADHD are also more distracted by them? The current data do not strongly support this claim, as the two event-related measures that (based on past research) are most directly associated with distractibility/attention capture, were not observed here at all – the P300 ERP response and event-related gaze-shifts.

Surprising or distracting stimuli often attract overt-attention, which manifests in stimulus-triggered gaze-shifts (Parkhurst et al., 2002; Geisler and Cormack, 2011; Nissens et al., 2016). Previous VR classroom studies, using embedded eye-tracking, showed that surprising visual stimuli – such objects appearing in the window – did elicit event-related gaze-shifts, and more so in children with ADHD (Stokes et al., 2022). However, the background sound-events in the current setup, which were only auditory and relatively contextual, did not elicit gaze-shifts towards the spatial location of the sounds (or any other location. This is actually in-line with previous findings by Brown et al. (2023) who also showed that hearing semantically-salient words in background speech (e.g., your name or semantic violations) elicited neural and SC responses, but was not

accompanied by gaze-shifts. This may suggest that auditory stimuli are, generally, less effective at capturing overt attention (Turoman and Vergauwe, 2023; Mandal et al., 2024), or that more salient auditory stimuli would be required to prompt listeners to move their gaze in search of the source of the disturbance.

Of course, irrelevant stimuli can capture attention covertly as well, and this is often associated with a family P300-like later neural responses, and specifically the P3a ERP component. The P3a is typically observed at a latency of ~250-400ms after hearing novel or surprising sounds (SanMiguel et al., 2010; Wetzel et al., 2013; Masson and Bidet-Caulet, 2019; Barry et al., 2020), and we had assumed that the background sound-events in our VR classroom would also elicit this response. Moreover, past studies have shown reduced P3a responses in individuals with ADHD (Kaiser et al., 2020), leading us to expect a similar pattern here. However, the ERPs measured here for background sounds are primarily dominated by the early sensory components, and show little evidence for later responses, in either group. This discrepancy may be explained by differences in the 'functional status' of background sound-events and task requirements of the current VR classroom relative to more traditional paradigms. The P3a is typically observed in tasks that require high levels of perceptual vigilance to discriminate between specific stimulus features, such as the pitch, duration or intensity of sounds presented in an oddball paradigm. In these tasks, so-called 'novel' or 'distractor' sounds/stimuli are presented as part of the sequence of stimuli that need to be discriminated, and these stimuli elicit a P3a [whereas target sounds elicit a somewhat similar P3b response; (Barry et al., 2020)]. Accordingly, although these 'distractor' stimuli do not require an explicit response, they are in fact within the focus of participants' endogenous attention, rather than truly 'background' stimuli (see review by Makov et al. 2022). Moreover, the magnitude of P3 responses is affected by factors such as the similarity between target and novel stimuli, their relative frequency and task requirements, suggesting an active comparative process with target stimuli. This differs substantially from the role of the novel sounds-events in the VR Classroom task used here, where sound-events are clearly in the background– both spatially and semantically - and occur in parallel to an ongoing speech stimulus (the lecture) that participants are primarily paying attention to. In addition, the perceptual and cognitive demands of the VR Classroom – where participants are instructed to listen and understand the speech as they would in a real-life classroom - differ substantially from the highly vigilant demands of speeded responses tasks that require perceptual judgment of every individual sound. These important differences in task and

stimulus characteristic ultimately lead to a different treatment of novel sound-events by the brain, and suggest that in the VR classroom these sounds do not 'capture attention' in the same way as novel sounds do in oddball-type paradigms. This comparison also highlights that the degree to which non-target stimuli are processes depends highly on contextual and cognitive factors, and many vary substantially across experimental designs (Makov et al., 2023; Mandal et al., 2024).

Sustained Attention to VR Classroom Learning

Besides the potential momentary distraction by sound-events, another key hypothesis we wanted to test here was that individuals with ADHD generally have a more difficult time maintaining attention to a frontal lesson in a classroom-setting. This was expected to manifest in reduced performance, reduced neural speech tracking, altered levels of arousal, heightened alpha and beta power and increased frequency of spontaneous gaze-shifts. The planned between-group comparisons did not show any of these patterns, although our exploratory multiple regression analysis did show that alpha-power and the frequency of spontaneous gaze-shifts carried significant explanatory power of ASRS scores. These results indicate that – counter to some prevalent notions of ADHD – this group is not "fundamentally flawed" on all aspects of attention, and are often indistinguishable from their controlled-peers on many neurophysiological metrics, particularly when tested under ecological circumstances (Barkley, 2019; Faraone et al., 2021; Sonuga-Barke et al., 2023). Below we discuss the lack of systematic group-differences in the current data set relative to more traditional attention-research, and the implications of these results for furthering a more dimensional perspective of attention and its real-life deficits (Heidbreder, 2015; Morris et al., 2022).

We start with discussing behavioral performance, which was similar for the two groups in both the Quiet and Events condition. The fact that individuals with ADHD were able to comprehend the mini-lectures and answer questions just as well as their peers indicates that they sustained attention to the teacher sufficiently and were not severely distracted - either due to sound-events or mind-wandering – to the point of impaired performance. The behavioral task used here is, admittedly, less demanding than the speeded-response tasks used in traditional attention research, which often do differentiate between individuals with and without ADHD. However, those tasks are not representative of most real-life contexts that require attention, and create relatively 'extreme' conditions – encouraging failures of performance (Hall et al., 2016; Barkley,

2019; Arrondo et al., 2023). By using a realistic classroom-learning task that is more representative of the level of information processing required in real-life contexts, the current data shows that individuals – both with and without ADHD – are able to maintain good performance despite occasional acoustic disturbances(see also Mandal et al., 2024). Obviously, between-group differences might emerge if background disturbances had been more severe or learning-material was more demanding, but these results stress the importance of exploring the parameters affecting behavioral performance under realistic task contexts.

A similar pattern emerges when looking at the neural tracking response of the teacher's speech. Numerous studies have shown that the components of TRF to continuous speech is modulated by selective attention, and are smaller when speech is not attended (Ding and Simon, 2012b; Mesgarani and Chang, 2012; Zion Golumbic et al., 2013; Petersen et al., 2017; Vanthornhout et al., 2019; Kaufman and Zion-Golumbic, 2023). Therefore, if individuals with ADHD have a more difficult time paying attention to the teacher – due either to external distraction or to mind wandering – we would have expected them to show reduced TRFs or that TRFs would not explain the recorded neural activity as well as in the control group. However, this was not the case, as both groups showed highly similar speech-tracking responses (TRF) and levels of predictive power. Qualitatively, the TRF responses in the Events condition were also similar to those estimated in the Quiet condition, further suggesting that the existence of sound-events did not necessarily detract from neural encoding of the teacher's speech. That said, in the current data there was a discrepancy between the results of the encoding and decoding TRF models, with a small but significant between-group difference observed in the reconstruction accuracy of the speech in the decoding model. Although encoding and decoding TRF models often converge in their results, they are also differently affected by the level of noise in the data and other factors, which sometimes can lead to different outcomes (Crosse et al., 2021). For full transparency, we report both findings here and encourage additional follow-up research to look more deeply into this matter.

Two additional metrics that are associated with reduced sustained attention and were hypothesized to be increased in the ADHD group are level of alpha power and the frequency of spontaneous gaze-shifts. Higher alpha power is associated with reduced levels of attention/arousal, and alpha power is consistently found to increase in conditions of boredom, tiredness or prolonged time on tasks (Clarke et al., 2001; Dockree et al., 2007; Palva and Palva, 2007; Wöstmann et al.,

2017; Boudewyn and Carter, 2018; Jin et al., 2019; Haro et al., 2022). Several studies have also found higher baseline alpha power in individuals with ADHD vs. controls (Barry et al., 2003; Johnstone et al., 2013; Bozhilova et al., 2022; Michelini et al., 2022), however results are not always consistent (Loo and Makeig, 2012; Johnstone et al., 2013). Similarly, gaze-shifts are often used as a proxy for distraction, particularly under naturalistic conditions ('t Hart et al., 2009; Foulsham et al., 2011; Risko and Kingstone, 2011; Risko et al., 2012; Hoppe et al., 2018), and recent VR and real-life eye-tracking studies have found increased frequency and prolonged duration of gaze-shifts towards irrelevant locations in individuals with ADHD (Braga et al., 2016; Mauriello et al., 2022; Stokes et al., 2022). Although neither of these metrics showed significant between-group differences in the current dataset, they did contribute to explaining variability in ASRS scores across the sample, suggesting some association with the severity of ADHD symptoms. Importantly, though, since in the current data there was no correlation between these metrics and behavioral performance or neural speech tracking of the lesson, there is no evidence that increased alpha power and/or increased scanning of the environment ultimately lead to reduced ability to process the content of the mini-lectures in our VR classroom. It is crucial for future research seeking objective 'biomarkers' associated with attention and its deficits to find reliable ways to link them to relevant aspects of real-life behavior, and not simply to self-reports of difficulties in attention, which can be prone to bias (e.g., recency/salience/generalization effects; (Rabiner et al., 2010; Toplak et al., 2013; Brevik et al., 2020).

Caveats

We would like to highlight two important caveats pertaining to the specific sample tested here, which should be considered when interpreting the current results. First, this study was conducted on adults, whereas the majority of ADHD research – and particularly all VR classroom studies - have focused on children (Pollak et al., 2009; Neguț et al., 2017; Schweitzer and Rizzo, 2022; Seesjärvi et al., 2022; Stokes et al., 2022). ADHD is primarily defined as a childhood condition and although it often persists into adulthood, with profound implications on professional and academics outcomes, how the nature of the deficit changes with age is still unknown (Barkley et al., 2008). Therefore, it is highly possible that some of the group-differences observed in previous studies with children in previous studies simply disappear or manifest differently in adulthood. Second, classification of the ADHD group was based on their prior clinical diagnosis, which likely

varied across participants in terms of diagnosis-approach and age of diagnosis. Although the two groups were distinguishable in their self-reported ASRS scores, a more reliable assessment of the severity of current ADHD symptoms might have been achieved if we had conducted a full clinical assessment of all participants, as per the acceptable clinical guidelines (APA, 2022). Third, the adult cohort tested here included primarily university students, who do not represent all adults but rather are a sub-group with relatively high cognitive and learning abilities, who are accustomed to sitting in lectures for long periods of time (Henrich et al., 2010). Hence, it is possible that the lack of group-differences in the current data set is because these individuals have learned to deal with distraction and have developed appropriate strategies, despite their ADHD, which allow them to thrive in academic settings. This caveat highlights the importance of considering sampling techniques in group-comparison studies, which sometimes limits their interpretation to more circumscribed sub-samples (Rad et al., 2018). At the same time, even if the current findings pertain only to this highly-functioning academic subgroup of individuals with ADHD, they imply an impressive ability to adapt and cope with the naturally occurring distractions of realistic classrooms. Specifically, they suggest that withstanding distraction in real-life contexts may be improved through training and good habit-forming, even in individuals with ADHD. We hope to follow up on this research in the future, to deepen our understanding of the diversity within the ADHD and typical population and to better characterize attention and distractibility across different types of real-life scenarios.

Synthesis and Conclusions

Recent years have seen a reckoning in the field of ADHD research, acknowledging that the heterogeneity of attentional challenges experienced by different individuals, and their frequent co-morbidity with other cognitive, neuropsychiatric or neurological conditions, makes it very difficult to distill clear-cut and objective behavioral, neural or physiological indicators of attention-disorders (Faraone et al., 2021; Sonuga-Barke et al., 2023). This difficulty is further exacerbated by the great diversity in diagnosis approaches, its largely subjective nature, and the fact that cognitive tools used for clinical assessments and research do not really capture the challenges faced in real life (Hall et al., 2016; Barkley, 2019; Mulraney et al., 2022; Arrondo et al., 2023; Schweitzer and Zion Golumbic, 2023). Here we offer a new approach for advancing the field of attention-research, by utilizing VR technology to simulate real-life experiences together with comprehensive

measurement of different neurophysiological responses, in attempt to get a more objective and well-rounded description of how individuals deal with the plethora of stimuli and task demands of their environments.

While our results point to some differences between individuals with and without a diagnosis of ADHD, they overwhelmingly demonstrate that most metrics associated with attention and information processing are actually similar across groups. These type of 'null results' are not always highlighted in research of clinical conditions (Joober et al., 2012; Franco et al., 2014), but are actually a critical component for developing objective metrics that reliably discern between 'neurotypical' levels of distraction/difficulties in attention vs. truly extreme cases, that may require clinical interventions. Moreover, the current results show that although the brain generates larger responses to irrelevant background sounds, and some individuals even show high levels of alpha-oscillations and frequent gaze-shifts indicative of distraction/boredom, this did not always detract from being able to perform the learning task at hand, suggesting a relatively good ability for multiplexed processing of concurrent stimuli, at least in the highly skilled student population tested here. While we acknowledge that many aspects of the current data require replication and generalization to more complex VR scenarios, we believe that this work helps advance the establishment of more objective and dimensional measures for capturing the different attention challenges experienced in real-life outside the lab.

Acknowledgements

This work was supported by the Israel Science Foundation (ISF grant # 2339/20 to EZG), the National Institute of Mental Health (NIMH grant R33MH110043 to JS) and the Binational Science Foundation (BSF grant # 2022024 to EZG & JS).

References

Anderson NC, Bischof WF, Kingstone A (2023) Eye tracking in virtual reality. :1–28.

APA (2022) Diagnosis and statistical manual of mental disorders, 5th ed. Arlington, VA: American Psychological Association.

Arrondo G, Mulraney M, Iturmendi-Sabater I, Musullulu H, Gambra L, Niculcea T, Banaschewski T, Simonoff E, Döpfner M, Hinshaw SP, Coghill D, Cortese S (2023) Systematic review and meta-analysis: clinical utility of continuous performance tests for the identification of attention-deficit/hyperactivity disorder. J Am Acad Child Adolesc Psychiatry 0.

Avisar A, Shalev L (2011) Sustained attention and behavioral characteristics associated with adhd in adults. Appl Neuropsychol 18:107–116.

Barkley RA (2019) Neuropsychological testing is not useful in the diagnosis of adhd: stop it (or prove it)! ADHD Rep 27:1–8.

Barkley RA, Murphy KR, Fischer M (2008) ADHD in adults : what the science says. New York, USA: Guilford Press.

Barry RJ, Johnstone SJ, Clarke AR (2003) A review of electrophysiology in attention-deficit/hyperactivity disorder: ii. event-related potentials. Clinical Neurophysiology 114:184–198.

Barry RJ, Steiner GZ, De Blasio FM, Fogarty JS, Karamacoska D, MacDonald B (2020) Components in the p300: don't forget the novelty p3! Psychophysiology 57:e13371.

Bellato A, Arora I, Hollis C, Groom MJ (2020) Is autonomic nervous system function atypical in attention deficit hyperactivity disorder (adhd)? a systematic review of the evidence. Neurosci Biobehav Rev 108:182–206.

Benedek M, Kaernbach C (2010) A continuous measure of phasic electrodermal activity. J Neurosci Methods 190:80–91.

Berger I, Cassuto H (2014) The effect of environmental distractors incorporation into a cpt on sustained attention and adhd diagnosis among adolescents. J Neurosci Methods 222:62–68.

Berti S (2013) The role of auditory transient and deviance processing in distraction of task performance: a combined behavioral and event-related brain potential study. Front Hum Neurosci 7:54615.

Bidet-Caulet A, Bottemanne L, Fonteneau C, Giard M-H, Bertrand O (2015) Brain dynamics of distractibility: interaction between top-down and bottom-up mechanisms of auditory attention. Brain Topogr 28:423–436.

Boudewyn MA, Carter CS (2018) I must have missed that: alpha-band oscillations track attention to spoken language. Neuropsychologia 117:148–155.

Bozhilova N, Kuntsi J, Rubia K, Asherson P, Michelini G (2022) Event-related brain dynamics during mind wandering in attention-deficit/hyperactivity disorder: an experience-sampling approach. Neuroimage Clin 35:103068.

Braga RM, Fu RZ, Seemungal BM, Wise RJS, Leech R (2016) Eye movements during auditory attention predict individual differences in dorsal attention network activity. Front Hum Neurosci 10:164.

Brevik EJ, Lundervold AJ, Haavik J, Posserud MB (2020) Validity and accuracy of the adult attention-deficit/hyperactivity disorder (adhd) self-report scale (asrs) and the wender utah rating scale (wurs) symptom checklists in discriminating between adults with and without adhd. Brain Behav 10:e01605.

Brown A, Pinto D, Burgart K, Zvilichovsky Y, Zion-Golumbic E (2023) Neurophysiological evidence for semantic processing of irrelevant speech and own-name detection in a virtual café. Journal of Neuroscience 43:5045–5056.
Chennu S, Noreika V, Gueorguiev D, Blenkmann A, Kochen S, Ibáñez A, Owen AM, Bekinschtein TA (2013) Expectation and attention in hierarchical auditory prediction. Journal of Neuroscience 33:11194–11205.
Clarke AR, Barry RJ, McCarthy R, Selikowitz M (2001) Excess beta activity in children with attention-deficit/hyperactivity disorder: an atypical electrophysiological group. Psychiatry Res 103:205–218.
Coleman B, Marion S, Rizzo A, Turnbull J, Nolty A (2019) Virtual reality assessment of classroom – related attention: an ecologically relevant approach to evaluating the effectiveness of working memory training. Front Psychol 10:1851.
Crosse MJ, Zuk NJ, Di Liberto GM, Nidiffer AR, Molholm S, Lalor EC (2021) Linear modeling of neurophysiological responses to speech and other continuous stimuli: methodological considerations for applied research. Front Neurosci 15:705621.
Ding N, Simon JZ (2012a) Emergence of neural encoding of auditory objects while listening to competing speakers. Proceedings of the National Academy of Sciences 109:11854–11859.
Ding N, Simon JZ (2012b) Neural coding of continuous speech in auditory cortex during monaural and dichotic listening. J Neurophysiol 107:78–89.
Ding N, Simon JZ (2014) Cortical entrainment to continuous speech: functional roles and interpretations. Front Hum Neurosci 8:311.
Dockree PM, Kelly SP, Foxe JJ, Reilly RB, Robertson IH (2007) Optimal sustained attention is linked to the spectral content of background eeg activity: greater ongoing tonic alpha (approximately 10 hz) power supports successful phasic goal activation. Eur J Neurosci 25:900–907.
Donoghue T, Haller M, Peterson EJ, Varma P, Sebastian P, Gao R, Noto T, Lara AH, Wallis JD, Knight RT, Shestyuk A, Voytek B (2020) Parameterizing neural power spectra into periodic and aperiodic components. Nat Neurosci 23:1655–1665.
Elahi H, Iosif AM, Mukherjee P, Hinshaw SP, Schweitzer JB (2023) Using hot and cool measures to phenotype and predict functional outcomes across dimensions of adhd and typical development in adolescents. Res Child Adolesc Psychopathol:1–15.
Escera C, Alho K, Winkler I, Näätänen R (1998) Neural mechanisms of involuntary attention to acoustic novelty and change. J Cogn Neurosci 10:590–604.
Escera C, Yago E, Corral M-J, Corbera S, Nuñez MI (2003) Attention capture by auditory significant stimuli: semantic analysis follows attention switching. European Journal of Neuroscience 18:2408–2412.
Esterman M, Rothlein D (2019) Models of sustained attention. Curr Opin Psychol 29:174–180.
Faraone S V. et al. (2021) The world federation of adhd international consensus statement: 208 evidence-based conclusions about the disorder. Neurosci Biobehav Rev 128:789–818.
Foulsham T, Walker E, Kingstone A (2011) The where, what and when of gaze allocation in the lab and the natural environment. Vision Res 51:1920–1931.
Franco A, Malhotra N, Simonovits G (2014) Publication bias in the social sciences: unlocking the file drawer. Science (1979) 345:1502–1505.
Frith CD, Allen HA (1983) The skin conductance orienting response as an index of attention. Biol Psychol 17:27–39.
Geisler WS, Cormack LK (2011) Models of overt attention. Oxford University Press.

Gloss D, Varma JK, Pringsheim T, Nuwer MR (2016) Practice advisory: the utility of eeg theta/beta power ratio in adhd diagnosis. Neurology 87:2375–2379.
Gray SA, Dueck K, Rogers M, Tannock R (2017) Qualitative review synthesis: the relationship between inattention and academic achievement. Educational Research 59:17–35.
Grosbras MH, Laird AR, Paus T (2005) Cortical regions involved in eye movements, shifts of attention, and gaze perception. In: Human Brain Mapping, pp 140–154. Wiley-Blackwell.
Gumenyuk V, Korzyukov O, Escera C, Hämäläinen M, Huotilainen M, Häyrinen T, Oksanen H, Näätänen R, Von Wendt L, Alho K (2005) Electrophysiological evidence of enhanced distractibility in adhd children. Neurosci Lett 374:212–217.
Hall CL, Valentine AZ, Groom MJ, Walker GM, Sayal K, Daley D, Hollis C (2016) The clinical utility of the continuous performance test and objective measures of activity for diagnosing and monitoring adhd in children: a systematic review. Eur Child Adolesc Psychiatry 25:677–699.
Haro S, Rao HM, Quatieria TF, Smalt CJ (2022) EEG alpha and pupil diameter reflect endogenous auditory attention switching and listening effort. European Journal of Neuroscience.
Heidbreder R (2015) ADHD symptomatology is best conceptualized as a spectrum: a dimensional versus unitary approach to diagnosis. ADHD Attention Deficit and Hyperactivity Disorders 2015 7:4 7:249–269.
Heil P (1997) Auditory cortical onset responses revisited. ii. response strength. J Neurophysiol 77:2642–2660.
Henrich J, Heine SJ, Norenzayan A (2010) The weirdest people in the world? Behavioral and Brain Sciences 33:61–83.
Hillyard S, Hink R, Schwent V, Picton T (1973) Electrical signs of selective attention in the human brain. Science (1979) 182:177–180.
Hobbiss MH, Lavie N (2024) Sustained selective attention in adolescence: cognitive development and predictors of distractibility at school. J Exp Child Psychol 238:105784.
Holtze B, Jaeger M, Debener S, Adiloğlu K, Mirkovic B (2021) Are they calling my name? attention capture is reflected in the neural tracking of attended and ignored speech. Front Neurosci 15.
Hoppe S, Loetscher T, Morey SA, Bulling A (2018) Eye movements during everyday behavior predict personality traits. Front Hum Neurosci 12:105.
Huang N, Elhilali M (2020) Push-pull competition between bottom-up and top-down auditory attention to natural soundscapes. Elife 9.
Jin CY, Borst JP, Vugt MK Van (2019) Predicting task-general mind-wandering with eeg. :1059–1073.
Johnstone SJ, Barry RJ, Clarke AR (2013) Ten years on: a follow-up review of erp research in attention-deficit/hyperactivity disorder. Clinical Neurophysiology 124:644–657.
Joober R, Schmitz N, Annable L, Boksa P (2012) Publication bias: what are the challenges and can they be overcome? Journal of Psychiatry and Neuroscience 37:149–152.
Kaiser A, Aggensteiner PM, Baumeister S, Holz NE, Banaschewski T, Brandeis D (2020) Earlier versus later cognitive event-related potentials (erps) in attention-deficit/hyperactivity disorder (adhd): a meta-analysis. Neurosci Biobehav Rev 112:117–134.
Kaufman M, Zion-Golumbic E (2023) Listening to two speakers: capacity and tradeoffs in neural speech tracking during selective and distributed attention. Neuroimage:119984.

Lalor EC, Foxe JJ (2010) Neural responses to uninterrupted natural speech can be extracted with precise temporal resolution. European Journal of Neuroscience 31:189–193.
Lange K (2013) The ups and downs of temporal orienting: a review of auditory temporal orienting studies and a model associating the heterogeneous findings on the auditory n1 with opposite effects of attention and prediction. Front Hum Neurosci 7.
Lim CL, Barry RJ, Gordon E, Sawant A, Rennie C, Yiannikas C (1996) The relationship between quantified eeg and skin conductance level. International Journal of Psychophysiology 21:151–162.
Litner B (2003) Teens with adhd: the challenge of high school. Child Youth Care Forum 32:137–158.
Loe IM, Feldman HM (2007) Academic and educational outcomes of children with adhd. J Pediatr Psychol 32:643–654.
Loo SK, Makeig S (2012) Clinical utility of eeg in attention-deficit/hyperactivity disorder: a research update. Neurotherapeutics 9:569–587.
Lütkenhöner B, Klein JS (2007) Auditory evoked field at threshold. Hear Res 228:188–200.
MacDonald B, Barry RJ (2020) Integration of three investigations of novelty, intensity, and significance in dishabituation paradigms: a study of the phasic orienting reflex. International Journal of Psychophysiology 147:113–127.
Makov S, Pinto D, Har-shai Yahav P, Miller LM, Zion Golumbic E (2023) "Unattended, distracting or irrelevant": theoretical implications of terminological choices in auditory selective attention research. Cognition 231.
Mandal A, Liesefeld AM, Liesefeld HR (2024) The surprising robustness of visual search against concurrent auditory distraction. J Exp Psychol Hum Percept Perform 50:99–118.
Marcus DK, Barry TD (2011) Does attention-deficit/hyperactivity disorder have a dimensional latent structure? a taxometric analysis. J Abnorm Psychol 120:427–442.
Masson R, Bidet-Caulet A (2019) Fronto-central p3a to distracting sounds: an index of their arousing properties. Neuroimage 185:164–180.
Mauriello C, Pham E, Kumar S, Piguet C, Deiber M-P, Aubry J-M, Dayer A, Michel CM, Perroud N, Berchio C (2022) Dysfunctional temporal stages of eye-gaze perception in adults with adhd: a high-density eeg study. Biol Psychol 171:108351.
Merrill BM, Raiker JS, Mattfeld AT, Macphee FL, Ramos MC, Zhao X, Altszuler AR, Schooler JW, Coxe S, Gnagy EM, Greiner AR, Coles EK, Pelham WE (2022) Mind-wandering and childhood adhd: experimental manipulations across laboratory and naturalistic settings. Res Child Adolesc Psychopathol 50:1139–1149.
Mesgarani N, Chang EF (2012) Selective cortical representation of attended speaker in multi-talker speech perception. Nature 485:233–236.
Michelini G, Salmastyan G, Vera JD, Lenartowicz A (2022) Event-related brain oscillations in attention-deficit/hyperactivity disorder (adhd): a systematic review and meta-analysis. International Journal of Psychophysiology 174:29–42.
Morris SE, Sanislow CA, Pacheco J, Vaidyanathan U, Gordon JA, Cuthbert BN (2022) Revisiting the seven pillars of rdoc. BMC Med 20:220.
Mulraney M, Arrondo G, Musullulu H, Iturmendi-Sabater I, Cortese S, Westwood SJ, Donno F, Banaschewski T, Simonoff E, Zuddas A, Döpfner M, Hinshaw SP, Coghill D (2022) Systematic review and meta-analysis: screening tools for attention-deficit/hyperactivity disorder in children and adolescents. J Am Acad Child Adolesc Psychiatry 61:982–996.

Murray AL, Booth T, Ribeaud D, Eisner M (2018) Disagreeing about development: an analysis of parent-teacher agreement in adhd symptom trajectories across the elementary school years. Int J Methods Psychiatr Res 27:e1723.
Näätänen R, Picton T (1987) The n1 wave of the human electric and magnetic response to sound: a review and an analysis of the component structure. Psychophysiology 24:375–425.
Narad ME, Garner AA, Peugh JL, Tamm L, Antonini TN, Kingery KM, Simon JO, Epstein JN (2015) Parent-teacher agreement on adhd symptoms across development. Psychol Assess 27:239–248.
Neguț A, Jurma AM, David D (2017) Virtual-reality-based attention assessment of adhd: clinicavr: classroom-cpt versus a traditional continuous performance test. Child Neuropsychology 23:692–712.
Nissens T, Failing M, Theeuwes J (2016) People look at the object they fear: oculomotor capture by stimuli that signal threat. https://doi.org/101080/0269993120161248905 31:1707–1714.
Palva S, Palva JM (2007) New vistas for α-frequency band oscillations. Trends Neurosci 30:150–158.
Parkhurst D, Law K, Niebur E (2002) Modeling the role of salience in the allocation of overt visual attention. Vision Res 42:107–123.
Parsons TD (2015) Virtual reality for enhanced ecological validity and experimental control in the clinical, affective and social neurosciences. Front Hum Neurosci 9:1–19.
Parsons TD, Duffield T, Asbee J (2019) A comparison of virtual reality classroom continuous performance tests to traditional continuous performance tests in delineating adhd: a meta-analysis. Neuropsychol Rev 29:338–356.
Peisch V, Rutter T, Wilkinson CL, Arnett AB (2021) Sensory processing and p300 event-related potential correlates of stimulant response in children with attention-deficit/hyperactivity disorder: a critical review. Clinical Neurophysiology 132:953–966.
Petersen EB, Wöstmann M, Obleser J, Lunner T (2017) Neural tracking of attended versus ignored speech is differentially affected by hearing loss. J Neurophysiol 117:18–27.
Pollak Y, Weiss PL, Rizzo AA, Weizer M, Shriki L, Shalev RS, Gross-Tsur V (2009) The utility of a continuous performance test embedded in virtual reality in measuring adhd-related deficits. J Dev Behav Pediatr 30:2–6.
Posner MI (1980) Orienting of attention. Q J Exp Psychol 32:3–25.
Rabiner DL, Murray DW, Rosen L, Hardy K, Skinner A, Underwood M (2010) Instability in teacher ratings of children's inattentive symptoms. Journal of Developmental & Behavioral Pediatrics 31:175–180.
Rad MS, Martingano AJ, Ginges J (2018) Toward a psychology of homo sapiens: making psychological science more representative of the human population. Proc Natl Acad Sci U S A 115:11401–11405.
Risko EF, Anderson N, Sarwal A, Engelhardt M, Kingstone A (2012) Everyday attention: variation in mind wandering and memory in a lecture. Appl Cogn Psychol 26:234–242.
Risko EF, Kingstone A (2011) Eyes wide shut: implied social presence, eye tracking and attention. Atten Percept Psychophys 73:291–296.
Rizzo ASA, Bowerly T, Buckwalter JG, Klimchuk D, Mitura R, Parsons TD, Bowerly T, Buckwalter JG, Rizzo ASA, Adams R, Finn P, Moes E, Flannery K, Rizzo ASA (2007) A virtual reality scenario for all seasons: the virtual classroom. Child Neuropsychol 13:35–44.
SanMiguel I, Morgan HM, Klein C, Linden D, Escera C (2010) On the functional significance of novelty-p3: facilitation by unexpected novel sounds. Biol Psychol 83:143–152.

Schomaker J, Walper D, Wittmann BC, Einhäuser W (2017) Attention in natural scenes: affective-motivational factors guide gaze independently of visual salience. Vision Res 133:161–175.
Schwartze M, Kotz SA (2015) The timing of regular sequences: production, perception, and covariation. J Cogn Neurosci 27:1697–1707.
Schweitzer JB, Rizzo A "Skip" (2022) Virtual reality and adhd: clinical assessment and treatment in the metaverse. https://doi.org/101521/adhd20223031 30:1–9.
Schweitzer JB, Zion Golumbic E (2023) Editorial: the use of continuous performance tasks in attention-deficit/hyperactivity disorder diagnosis: a cautionary note. J Am Acad Child Adolesc Psychiatry 0.
Seesjärvi E, Puhakka J, Aronen ET, Lipsanen J, Mannerkoski M, Hering A, Zuber S, Kliegel M, Laine M, Salmi J (2022) Quantifying adhd symptoms in open-ended everyday life contexts with a new virtual reality task. J Atten Disord 26:1394–1411.
Selaskowski B, Asché LM, Wiebe A, Kannen K, Aslan B, Gerding TM, Sanchez D, Ettinger U, Kölle M, Lux S, Philipsen A, Braun N (2023) Gaze-based attention refocusing training in virtual reality for adult attention-deficit/hyperactivity disorder. BMC Psychiatry 23:74.
Sergeant J (2000) The cognitive-energetic model: an empirical approach to attention-deficit hyperactivity disorder. In: Neuroscience and Biobehavioral Reviews, pp 7–12. Pergamon.
Smallwood J, Fishman DJ, Schooler JW (2007) Counting the cost of an absent mind: mind wandering as an underrecognized influence on educational performance. Psychon Bull Rev 14:230–236.
Sonuga-Barke EJS, Becker SP, Bölte S, Castellanos FX, Franke B, Newcorn JH, Nigg JT, Rohde LA, Simonoff E (2023) Annual research review: perspectives on progress in adhd science – from characterization to cause. Journal of Child Psychology and Psychiatry 64:506–532.
Stokes JD, Rizzo A, Geng JJ, Schweitzer JB (2022) Measuring attentional distraction in children with adhd using virtual reality technology with eye-tracking. Front Virtual Real 3.
Szpunar KK, Moulton ST, Schacter DL (2013) Mind wandering and education: from the classroom to online learning. Front Psychol 4:495.
't Hart BM, Vockeroth J, Schumann F, Bartl K, Schneider E, König P, Einhäuser W (2009) Gaze allocation in natural stimuli: comparing free exploration to head-fixed viewing conditions. Vis cogn 17:1132–1158.
Thomson DR, Besner D, Smilek D (2015) A resource-control account of sustained attention. Perspectives on Psychological Science 10:82–96.
Toplak ME, West RF, Stanovich KE (2013) Practitioner review: do performance-based measures and ratings of executive function assess the same construct? Journal of Child Psychology and Psychiatry 54:131–143.
Turoman N, Vergauwe E (2023) The effect of multisensory distraction on working memory: a role for task relevance? J Exp Psychol Learn Mem Cogn.
Vanthornhout J, Decruy L, Francart T (2019) Effect of task and attention on neural tracking of speech. Front Neurosci 13.
Wetzel N, Schröger E, Widmann A (2013) The dissociation between the p3a event-related potential and behavioral distraction. Psychophysiology 50:920–930.
Woldorff MG, Gallen CC, Hampson SA, Hillyard SA, Pantev C, Sobel D, Bloom FE (1993) Modulation of early sensory processing in human auditory cortex during auditory selective attention. Proceedings of the National Academy of Sciences 90:8722–8726.

Woldorff MG, Hillyard SA (1991) Modulation of early auditory processing during selective listening to rapidly presented tones. Electroencephalogr Clin Neurophysiol 79:170–191.
Wöstmann M, Lim S-J, Obleser J (2017) The human neural alpha response to speech is a proxy of attentional control. Cerebral Cortex 27:3307–3317.
Zimmer H, Richter F (2023) Novelty detection and orienting: effects on skin conductance and heart rate. Psychol Res 87:1101–1113.
Zion Golumbic EM, Ding N, Bickel S, Lakatos P, Schevon CA, McKhann G, Goodman RR, Emerson R, Mehta AD, Simon JZ, Poeppel D, Schroeder CE (2013) Mechanisms underlying selective neuronal tracking of attended speech at a cocktail party . Neuron 77:980–991.