Selective attention and sensitivity to auditory disturbances in a virtually-real Classroom: Comparison of adults with and without AD(H)D

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 individuals with Attention Deficit (Hyperactivity) Disorder (AD(H)D), find it difficult to maintain attention during classroom learning. However, traditional paradigms used to evaluate attention do not capture the complexity and dynamic nature of real-life classrooms. Using a novel Virtual Reality platform, coupled with measurement of neural activity, eye-gaze and skin conductance, here we studied the neurophysiological manifestations of attention and distractibility, under realistic learning conditions. Individuals with AD(H)D exhibited higher neural responses to irrelevant sounds and reduced speech tracking of the teacher, relative to controls. Additional neurophysiological measures, such the power of alpha-oscillations and frequency of gaze-shifts away from the teacher, contributed to explaining variance in self-reported AD(H)D symptoms across the sample. These ecologically-valid findings provide critical insight into the neurophysiological mechanisms underlying individual differences in the capacity for sustained attention and the proneness to distraction and mind-wandering, experienced in real-life situations.

Introduction

Many people find it difficult to maintain attention to a frontal classroom lecture. Doing so 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 Disorder or Attention-Deficit and Hyperactivity Disorder (referred to jointly as: AD(H)D), 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 AD(H)D 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 different types of ecological sound-events (non-verbal human sounds such as coughs and throat clearings; and artificial sounds such as phone ringtones) (see also Levy et al., 2024). 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 AD(H)D when they engage in realistic classroom learning, and which metrics reliably predict the severity of AD(H)D-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 captures aspects of its sensory and 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 AD(H)D and/or in conditions that contain external disturbances; (2) Neural Event-Related Potentials (ERPs), transient changes in Skin Conductance (SC) and overt gaze-shifts following unexpected sound-events in the background of the classroom. These metrics are thought to reflect 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); (3) The frequency of gaze-shifts away from the teacher, and time spent looking around the classroom, metrics associated with attention-shifts and distractibility(Grosbras et al., 2005; Schomaker et al., 2017), and are often heightened in individuals with AD(H)D (Mauriello et al., 2022; Stokes et al., 2022; Selaskowski et al., 2023); (4) The power of alpha and beta oscillations, which are often associated with

107 increased mind-wandering or boredom (Clarke et al., 2001; Boudewyn and Carter, 2018; Jin et al.,
108 2019), and some have suggested may be altered in individuals with AD(H)D (although use of
109 spectral signatures as biomarkers for AD(H)D is highly controversial(Gloss et al., 2016)); (5)
110 Continuous levels of arousal, as measured by skin-conductance (SC), which some propose are
111 either heightened or reduced in individuals with AD(H)D relative to their control peers (Sergeant,
112 2000; Bellato et al., 2020).
113 This novel ecological approach and the rich repertoire of neurophysiological metrics measured
114 here, afford unique insights into the mechanistic underpinnings of paying attention in class and the
115 disruptive effects of background sounds, in both the neurotypical and AD(H)D population.
116
117
118 **Methods**
119 **Participants**
120 Fifty-four participants initially participated in the study; however, the data from five participants
121 were excluded due to technical issues or their voluntary request to discontinue participation. The
122 final sample consisted of 49 participants (21 male, 28 female; 46 right-handed, 3 left-handed). Of
123 these participants, 24 had a prior clinical diagnosis of attention deficit hyperactivity disorder
124 (AD(H)D group), and 25 participants did not (control group). The two groups were matched for
125 age, which ranged between 20 and 38 years (mean age of all participants: 26.9 ± 3.0; AD(H)D
126 participants: 26.7 ± 3.7; control participants: 25.7 ± 4.4). The majority of participants were
127 university students. Participants in the AD(H)D group were questioned about their medication
128 intake regimen, and the experiment was scheduled at least 12 hours after their last dose of
129 medication. None of the participants (in either group) reported taking any other medications on a
130 regular basis. The experimental protocol was approved by the ethics committee at Bar-Ilan
131 University, and all participants provided written informed consent prior to their involvement in the
132 study and data collection procedures. Participants received financial compensation or course credit
133 in exchange for their participation.
134
135 **ASRS questionnaire**
136 All participants completed a self-report questionnaire assessing AD(H)D symptoms (ASRS-v1.1;
137 Adler et al., 2006, Hebrew version: Zohar and Konfortes, 2010), which has been shown to have

138 good psychometric properties high sensitivity for AD(H)D diagnosis (Kessler et al., 2005).
139 Participants in the AD(H)D group were instructed to complete the questionnaire based on their
140 experiences at a time when they are not taking medication. The questionnaire consists of 18 items,
141 which are rated on a 5-point Likert scale ranging from 0 (never) to 4 (very often). ASRS scores
142 for each participant were calculated as the total sum of marks across all items, as this has been
143 shown to yield the highest sensitivity to AD(H)D diagnosis (Zohar and Konfortes, 2010). We did
144 not try to categorize AD(H)D-subtypes (e.g., inattention, hyperactivity, combined) since these
145 subtypes are less reliable in adulthood. As expected, the group diagnosed with AD(H)D exhibited
146 significantly higher ASRS scores compared to the control group [$t(47) = 8.49$, $p < 0.001$; see Figure
147 1]. This outcome validates the group allocation and demonstrates the consistency between
148 AD(H)D diagnosis and self-reported AD(H)D symptoms.

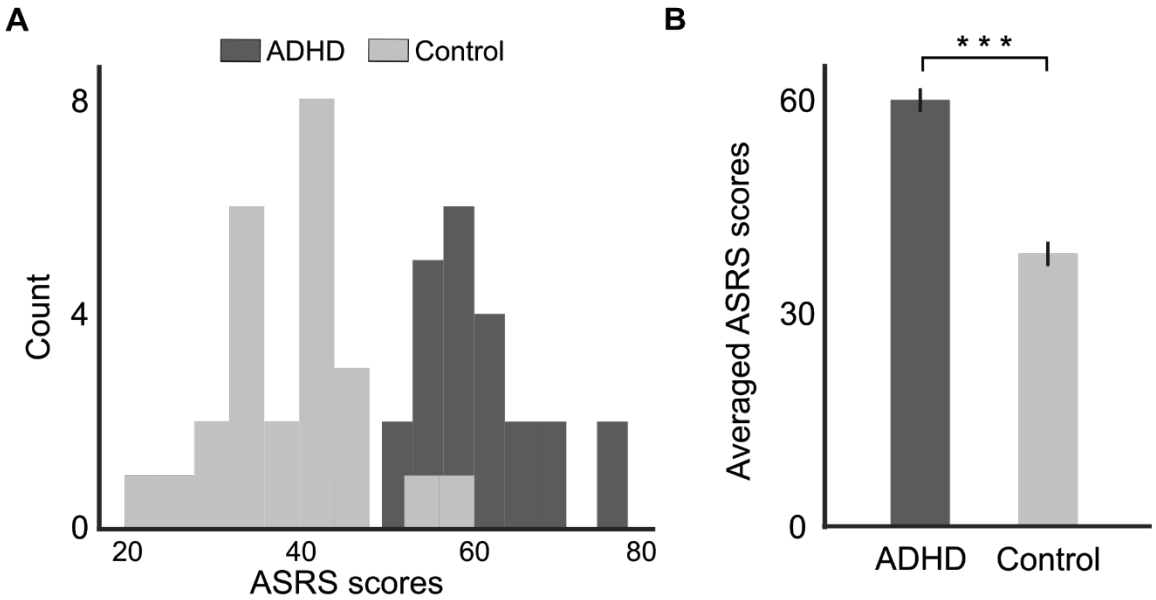


Figure 1. Histogram (A) and bar graph (B) showing the distribution of ASRS scores in the AD(H)D and Control groups. Error bars represent standard error of the mean (SEM). This confirms group-allocation, within the tested sample.

149

150 **The VR Classroom: Design and Stimuli**

151 The VR classroom platform was developed using the Unity game engine (unity.com; JavaScript
152 and C# programming). Development and programming of the VR classroom were done primarily
153 in-house, using assets (avatars and environment) were sourced from pre-existing databases. The
154 classroom environment was adapted from assets provided by Tirgames on TurboSquid

(https://www.turbosquid.com/Search/Artists/Tirgames) and modified to meet the experimental needs. The avatars and their basic animations were sourced from the Mixamo library, which at the time of development supported legacy avatars with facial blendshapes (this functionality is no longer available in current versions of Mixamo). A brief video example of the VR classroom is available at https://osf.io/rf6t8.

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 a range of scientific and 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. To achieve realistic lip-speech synchronization, the teacher's lip movements were controlled by the temporal envelope of the speech, adjusting both timing and mouth size dynamically. His body motions were animated using natural talking gestures.

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 Normalize function in the audio-editing software Audacity (theaudacityteam.org, ver 3.4), with the peak amplitude parameter set to -5 dB, and 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.

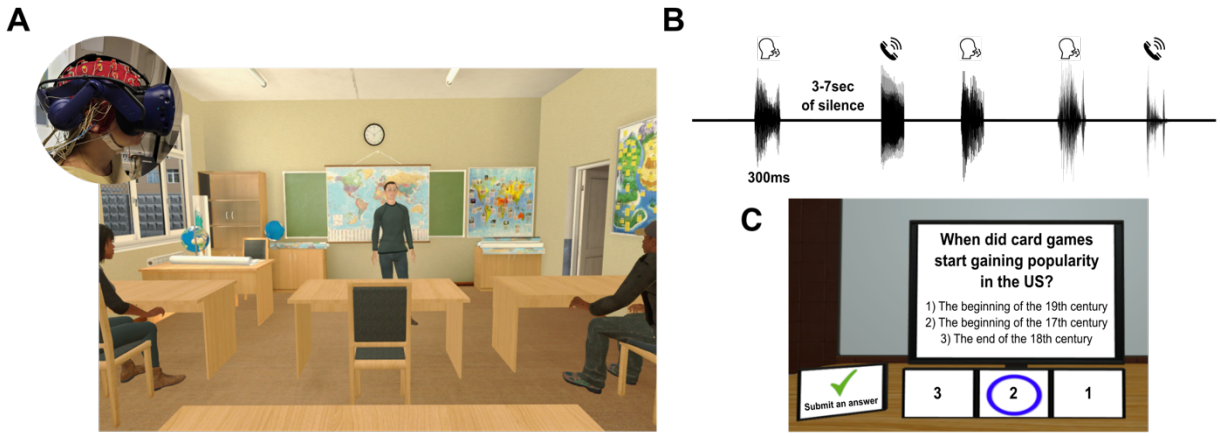


Figure 2. A. Illustration of the Virtual Classroom used in the study. Inset: a participant wearing the VR headset over the EEG cap. **B.** Example of the sequence of sound-events presented in Event-trials, which were emitted from a spatial location to the right or left of the participant. Event-trials 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 or rejecting the null hypothesis (H0; BF_{10} thresholds used throughout: $BF_{10} < 1$ indicates weak support of H0 and $BF_{10} < 0.25$ indicates moderate/strong support of H0; $BF_{10} > 1$ indicates weak support for rejecting H0 and $BF_{10} > 2.5$ indicates moderate/strong support for rejecting H0).

EEG data analysis

The EEG data was preprocessed using MATLAB with the FieldTrip toolbox (https://www.fieldtriptoolbox.org, version 20220729). The raw EEG data was referenced to linked right and left mastoids, then band-pass filtered between 0.5 and 40 Hz (4th order zero-phase Butterworth IIR filter), detrended and demeaned. Visual inspection was performed to identify and remove gross artifacts (excluding eye movements). Independent Component Analysis (ICA) was then used to further remove components associated with horizontal or vertical eye movements and heartbeats, which were selected manually based on their characteristic time-course patterns and topographical distributions, consistent with known artifact signatures. 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 (4th order zero-phase Butterworth IIR filter), and downsampled to 100 Hz for computational efficiency. To ensure comparability across participants and conditions, the EEG data were normalized using z-score transformation. 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. Speech-tracking analysis was limited only to the Events condition, since the Quiet condition contained substantially less data (8 trials; ~ 5 minutes of data in total), which was insufficient for obtaining a reliable speech-tracking response (predictive power non-significant vs. permutations of shuffled data, $p=0.43$). For speech-tracking in the Events condition we used a multivariate approach, and included the temporal envelopes of

both the teacher's speech and the sound-events as separate regressors (S) when modeling the neural response, so as to represent the entire soundscape they heard. For both regressors, we derived a broadband temporal envelope by filtering the audio through a narrow-band cochlear filterbank, extracting the narrowband envelopes using a Hilbert transform, and averaging across bands. The regressors were then 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 ensure better generalization to new data. We tested a range of ridge parameter values (λ's; between 10^{-2}-10^{4}) and 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 (for each λ) 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. We report results of the model with the λ the yielded the highest predictive power at the group-level (rather than selecting a different λ for each participant which can lead to incomparable TRF models across participants; see discussion in Kaufman & Zion Golumbic 2023).

The statistical reliability 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.

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). Here too, we used a ridge-regression regularization approach and used a leave-one-out cross-validation procedure to assess the quality of the reconstruction. We tested a range of ridge parameter values (λ's; between 10^{-2}-10^{4}) and 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 quality of the model (for each λ) is estimated as the Pearson's correlation between the reconstructed S of a left-out-trial vs, the actual S ("Reconstruction Correlation"). We report results of the model with the λ the yielded the highest predictive power at the group-level. The statistical significance of the Reconstruction Correlation was assessed using a permutation test in which 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. This yields a null-distribution of predictive power values that could be obtained by chance, and the real model is considered reliable if its predictive power falls within the top 5%^{tile} of the null-distribution.

Speech tracking analysis – using both encoding and decoding approaches – was performed separately for each Group (AD(H)D and Control). To test for between-group differences, we compared the models' predictive power, using unpaired t-tests. We additionally calculated the Bayes Factor (BF_{10}) for each comparison, to test the level of confidence for embracing or rejecting the null hypothesis.

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 (4th order zero-phase Butterworth IIR filter), and baseline corrected to the pre-stimulus period (-100 to 0 ms). ERPs were derived separately for each group (AD(H)D 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 (AD(H)D 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 to test the level of confidence for embracing or rejecting the null hypothesis.

Skin Conductance data analysis

The Skin Conductance (SC) signal was analyzed using the Ledalab MATLAB toolbox (version 3.4.9; Benedek and Kaernbach, 2010; http://www.ledalab.de/) and custom-written scripts. The raw data was downsampled to 16Hz using FieldTrip's ft_resampledata function, which applies a built-in anti-aliasing low-pass filter to prevent aliasing artifacts. Data were inspected manually for any noticeable artifacts (large 'jumps'), and if present were corrected using linear interpolation in Ledalab. A continuous decomposition analysis (CDA) was employed to separate the tonic and phasic SC responses for each participant. The CDA was conducted using the 'sdeco' mode (signal decomposition), which iteratively optimizes the separation of tonic and phasic components using the default regularization settings. 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 (AD(H)D 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,

388 using the x, y, z coordinates in 3D space provided by the eye-tracker. 'Fixations' were defined as
389 clusters of adjacent time-points within a radius of less than 0.01 distance from each other, lasting
390 at least 80 ms. Each fixation was associated with a specific ROI in the VR classroom, indicating
391 the object in the virtual space that the participant looked at. 'Gaze-shifts' were defined as transitions
392 between two stable fixations.
393 To analyze spontaneous gaze-dynamics throughout an entire trial, we quantified the percent of
394 each trial that participants spent looking at the teacher (target) versus all other ROIs, as well as the
395 number of gaze-shifts away from the teacher per trial. Between-group differences in these
396 measures of spontaneous gaze patterns were assessed using unpaired t-tests, separately for the
397 Quiet and Events conditions. The Bayes Factor (BF_{10}) was calculated for each pairwise
398 comparison to test the level of confidence for embracing or rejecting the null hypothesis.
399
400 To determine whether background Sound-events elicited overt gaze shifts, we segmented the gaze-
401 data into 2-second long epochs around each Event-sound. As a control, we randomly chose a
402 similar number of 2-second long epochs from the Quiet-condition. We then counted how many
403 epochs included at least one gaze-shift away from the teacher, and used a paired t-test to determine
404 whether the frequency of gaze-shifts was higher following Sound-events relative to the Control
405 epochs (regardless of Group). We also used a Mixed ANOVA to test for potential differences in
406 the likelihood of gaze-shifts following the different event types (Artificial vs. Human non-verbal),
407 and for the two Groups (AD(H)D vs. Control). For this analysis, the proportion of gaze shifts was
408 normalized for each participant by subtracting the number of gaze shifts following Sound-events
409 relative to the control epochs.
410
411 **Multivariate Analyses**
412 Besides testing for effects of Group in each metric individually, we also performed multivariate
413 analyses to examine the relationship between the various metrics collected here and their ability
414 to predict ADHD diagnosis (binary) and/or the severity of AD(H)D symptoms.
415 This analysis focused on 10 key measures, based on the results of the univariate analyses (averaged
416 across conditions, when relevant): 5 measures related to the response to the teacher's speech:
417 Behavioral performance, neural speech tracking decoding (reconstruction correlation), the N1- and
418 P2 peaks of the speech tracking response (TRF-N1; TRF-P2), and the average number of gaze-

shifts away from the teacher; 3 measures related to responses to the sound events: Amplitude of the N1- and P2-responses to sound-events (ERP-N1; ERP-P2), phasic SC response to sound events (event-related SC); and 2 global measures associated with the listeners' 'state': alpha- and beta-power.

First, we calculated Spearman's correlation between all pairs of measures (fdr correction for multiple comparisons was applied). Next, we entered all factors into a multiple regression model and performed a dominance analysis to determine the relative importance of each measure for predicting whether an individual was in the ADHD or control group (logistic regression) or the severity of ADHD symptoms, as reflected by their ASRS scores (linear regression). Dominance analysis was performed as follows: For each of the 10 measures, we computed the difference in a model-fit index (AUC for logistic regression; R^2 for linear regression) between a full model containing all 10 predictors, and a model where the measure under consideration was held-out, as a means for assessing the unique contribution of each measure to overall model performance (dominance). We then sorted the measures based on their dominance and report the multiple regression model that explains the most amount of variance using the least number of variables. These analyses were conducted using the statistical software JASP (JASP Team, 2024).

Results

The current dataset is extremely rich, consisting of many different behavioral, neural and physiological responses. In reporting these results, we have separated between metrics that are associated with **paying attention to the teacher** (behavioral performance, neural tracking of the teacher's speech, and looking at the teacher), those capturing **responses to the irrelevant sound-events** (ERPs and event-related changes in SC and gaze); as well as more global neurophysiological measures that may be associated with the listeners' overall **'state' of attention or arousal** (alpha- and beta-power and tonic SC).

Paying attention to the teacher

First, we tested whether there are differences between individuals with and without AD(H)D in metrics associated with focusing ones' attention on the teacher - accuracy on answering comprehension questions, neural speech tracking of the teacher's speech and the gaze-patterns of focusing overt attention towards the teacher.

Due to the difference in number of trials statistical testing of between-group differences were performed separately for the Events and Quiet conditions (when applicable), and comparisons between conditions are evaluated only qualitatively.

455 Behavioral Accuracy

456 Participants demonstrated overall good performance on the comprehension task, achieving an
457 average accuracy of 87.87% (±6.77% SEM). This serves as verification that participants followed
458 the instructions, listened to teacher's speech and understood the content of the mini-lessons.
459 Performance levels were similarly good in both the Events and Quiet conditions, indicating that
460 the presence of occasional sound-events did not disrupt overall understanding of the mini-lessons.
461 No significant differences in performance were found between groups, in either condition [Events:
462 $t(47) = 1.052$, $p = 0.297$, Cohen's $d = 0.30$, $BF_{10} = 0.448$ (weak support of H0); Quiet: $t(47) =$
463 1.004, $p = 0.320$, Cohen's $d = 0.28$, $BF_{10} = 0.430$ (weak support for H0); Figure 3].

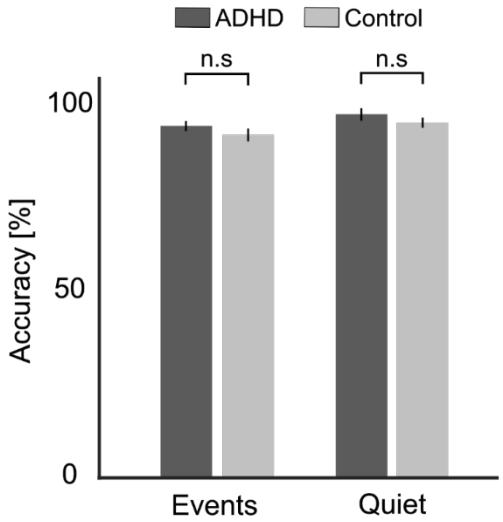


Figure 3. Accuracy on comprehension questions. Shown separately for the AD(H)D and Control groups, in the Quiet and Events conditions. Bar graphs represent average accuracy levels across participants and error bars represent the SEM. No significant differences were found between groups in either condition (n.s).

464

465 Speech tracking of teacher's speech

466 We conducted speech-tracking analysis of the neural response to the teacher's speech, using both
467 an encoding and decoding approach. As noted in the methods section, speech-tracking analysis
468 could only be performed reliably in the Events condition, and we used a multivariate approach to
469 account for variance in the neural activity due to the presence of event-sounds (see supplementary
470 materials for comparison with univariate analysis).

471 The TRFs estimated in response to the teachers' speech were similar across both groups showing
472 the typical negative peak at ~ 100ms (N1), followed by a positive peak ~200ms (P2), and the

model's predictive power following the traditional mid-central topography associated with auditory responses (Figure 4A,B).

No significant between-group differences were found in the predictive power of the encoding model (averaged across electrodes), and Bayes analysis indicates moderate support for the null-hypothesis [t(47) = 0.23, p = 0.814, Cohen's d = 0.067, BF_{10} = 0.15]. However, when using a decoding model, which considers all electrodes in one model, we did find slightly lower reconstruction correlations for the teacher's speech in the AD(H)D group relative to controls [t(47) = -1.948, p = 0.057, Cohen's d = -0.557, BF_{10} = 1.308 (weak/moderate support for H1); Figure 4C].

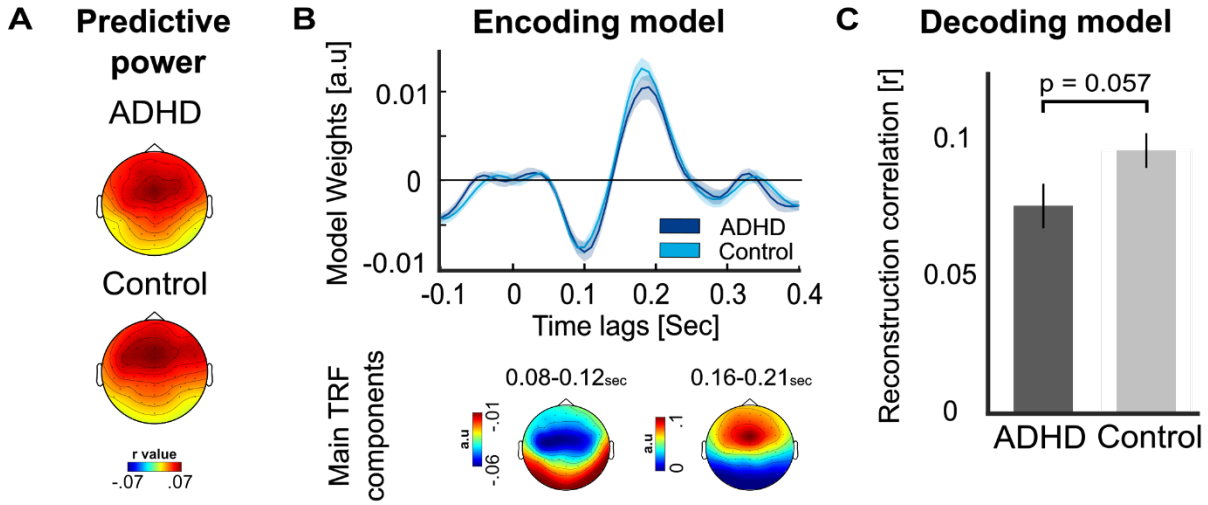


Figure 4. Speech-tracking of the teacher in the presence of sound-events. A. Topographical distribution of the predictive power values, estimated using the multivariate encoding model, in the ADHD and Control groups. No significant differences were observed between groups. B. TRFs estimated for each group in the Events condition, based on the average of electrodes FC1, FC2, and FCz, which exhibited the strongest activity for the two main components. Shaded areas represent the SEM. Below the TRFs, topographies of the main component are presented. C. Bar graphs showing the average reconstruction correlation of the decoding model (Pearson's r) in each group. The ADHD group showed slightly lower reconstruction correlations for the teacher's speech, though the effect did not reach statistical significance (p = 0.057). Error bars represent the SEM.

Looking at the teacher

Analysis of gaze-patterns showed 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 at were the blackboards behind the teacher, to the right and to the left (Figure 5A). 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 (AD(H)D vs. Control) separately in each Condition (Quiet and Events). No significant differences were found, however BF analysis indicated that the null-hypotheses was not strongly supported [*Percent gaze-time at teacher:* Events Condition: $t(47) = -0.899$, $p = 0.372$, Cohen's $d = -0.257$, $BF_{10} = 0.397$ (weak support for H0); Quiet Condition: $t(47) = -0.597$, $p = 0.553$, Cohen's $d = -0.170$, $BF_{10} = 0.33$ (weak support for H0) ; *Number of gaze-shifts away from teacher*: Events Condition*:* $t(47) = 1.265$, $p = 0.211$, Cohen's $d = 0.361$, $BF_{10} = 0.547$ (weak support for H0); Quiet Condition: $t(47) = 0.644$, $p = 0.522$, Cohen's $d = 0.184$, $BF_{10} = 0.338$ (weak support for H0) ; Figure 5C].

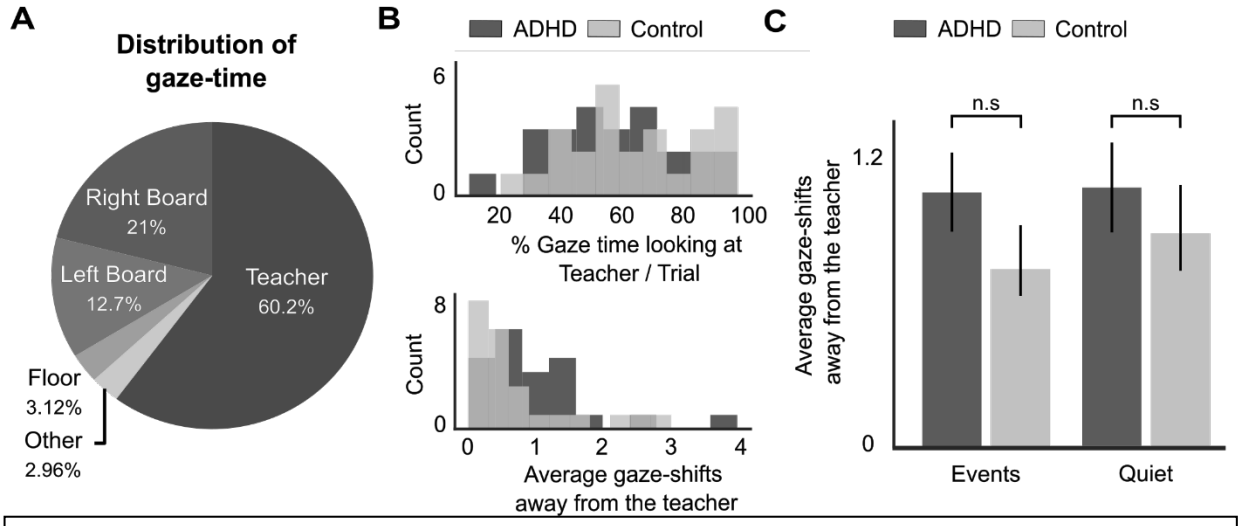


Figure 5. Eye-gaze results. A. Pie chart representing the average amount of time that participants spent looking at different areas within the VR classroom. **B.** Distribution of the proportion of gaze-time towards the teacher (top) and number of gaze-shifts performed away from the teacher (bottom), for all participants in the AD(H)D and Control groups. **C.** Bar graph represent the average number of gaze-shifts performed away from the teacher, shown separately for the AD(H)D and Control groups, and for the Quiet and Events conditions. No significant differences were found in any comparison (n.s)**.**

497 ***Responses to Irrelevant Sound-Events***

498 Next, we tested whether neurophysiological responses to the event-sounds themselves differed
499 between groups. The three event-related metrics tested here were: neural ERPs, transient changes
500 in skin-conductance (SC) and gaze-shifts following event-sounds. In addition, we compared the
501 response to the two different event-types presented here (Artificial vs. non-verbal Human
502 sounds).

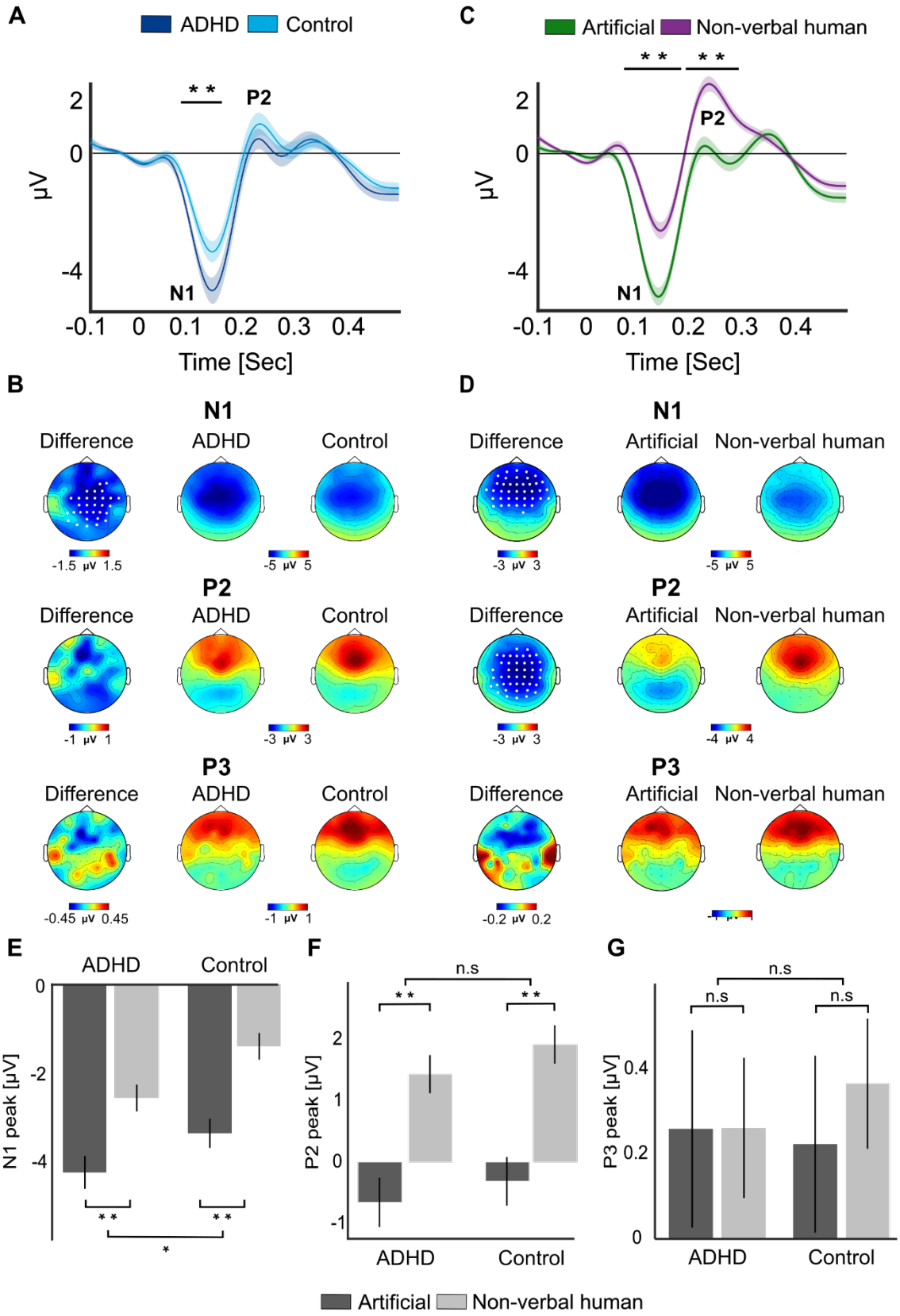
503

504 Neural Event Related Potentials (ERPs)

505 Visual inspection of the neural ERPs showed three prominent centro-parietal components - a
506 negative peak around 100ms, which corresponds to the early sensory N1; followed by two positive
507 peaks around 240ms and 350ms, which likely corresponds to the P2 and P3 responses, respectively
508 (Figure 6).

509

510 Statistical analysis comparing ERPs in the AD(H)D vs. Control groups (collapsed across both
511 Event-Types) showed that the early N1 response was significantly larger in the AD(H)D group
512 ($p<0.05$, cluster correction; Figure 6A&B). However, no significant differences between groups
513 were found for the later P2 response. When comparing ERPs to Artificial vs. non-verbal Human
514 sounds we found that both the N1 and the P2 were significantly modulated by Event-type, with a
515 larger N1 for the Artificial sounds and a larger P2 for the non-verbal Human sounds (both $p<0.002$,
516 cluster correction; Figure 6C&D). A mixed-ANOVA performed on the amplitudes of each
517 component confirmed these main effects, but did not reveal any significant interaction between
518 Group and Event-type [N1: *Group* [$F(1, 47) = 6.15$, $p = 0.044$, $\eta^2 = 0.11$], *Event type* [$F(1, 47) =$
519 63.82, $p < 10^{-9}$, $\eta^2 = 0.57$], *interaction* [$F(1, 47) = 0.389$, $p = 0.24$, $\eta^2 = 0.008$]; P2: *Group* [$F(1,$
520 $47) = 0.84$, $p = 0.36$, $\eta^2 = 0.017$], *Event type* [$F(1, 47) = 71.75$, $p < 10^{-9}$, $\eta^2 =0.6$], *interaction* [$F(1,$
521 $47) = 0.07$, $p=0.78$, $\eta^2 = 0.001$]; P3: *Group* [$F(1, 47) = 0.028$, $p = 0.876$, $\eta^2 = 0.0005$], *Event type*
522 [$F(1, 47) = 0.184$, $p = 0.669$, $\eta^2 = 0.003$], *interaction* [$F(1, 47) = 0.166$, $p = 0.684$, $\eta^2 = 0.003$];
523 Figure 6].


A
ADHD
Control
P2
N1
μV
Time [Sec]
C
Artificial
Non-verbal human
P2
N1
μV
Time [Sec]
B
N1
Difference
ADHD
Control
-1.5 μV 1.5
-5 μV 5
P2
Difference
ADHD
Control
-1 μV 1
-3 μV 3
P3
Difference
ADHD
Control
-0.45 μV 0.45
-1 μV 1
D
N1
Difference
Artificial
Non-verbal human
-3 μV 3
-5 μV 5
P2
Difference
Artificial
Non-verbal human
-3 μV 3
-4 μV 4
P3
Difference
Artificial
Non-verbal human
-0.2 μV 0.2
E
ADHD
Control
N1 peak [μV]
F
n.s
P2 peak [μV]
ADHD
Control
G
n.s
n.s
n.s
P3 peak [μV]
ADHD
Control
Artificial
Non-verbal human

524

Figure 6. Event Related Potentials in response to sound-events. A. Grand-average ERP to sound-events, shown separately for the AD(H)D 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 AD(H)D 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 indicate 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 (AD(H)D vs. Control) and Event-type (Artificial vs. Non-verbal human). ** $p < 0.001$; * $p<0.05$.

Event-Related increase in SC

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 7). This response is consistent with the well documented orienting-reflex, elicited following salient events.

A Mixed ANOVA showed a significant main effects of Event type, with a higher response following artificial sounds relative to non-verbal human sounds [$F(1, 47) = 7.88$, $p = 0.007$, $\eta^2 =$

534 0.143], however there was no significant main effect of Group [F(1, 47) = 0.4, p = 0.53, η^2 =
535 0.008] or interaction between them [F(1, 47) = 0.02, p = 0.87, η^2 = 0.0005; Figure 7].

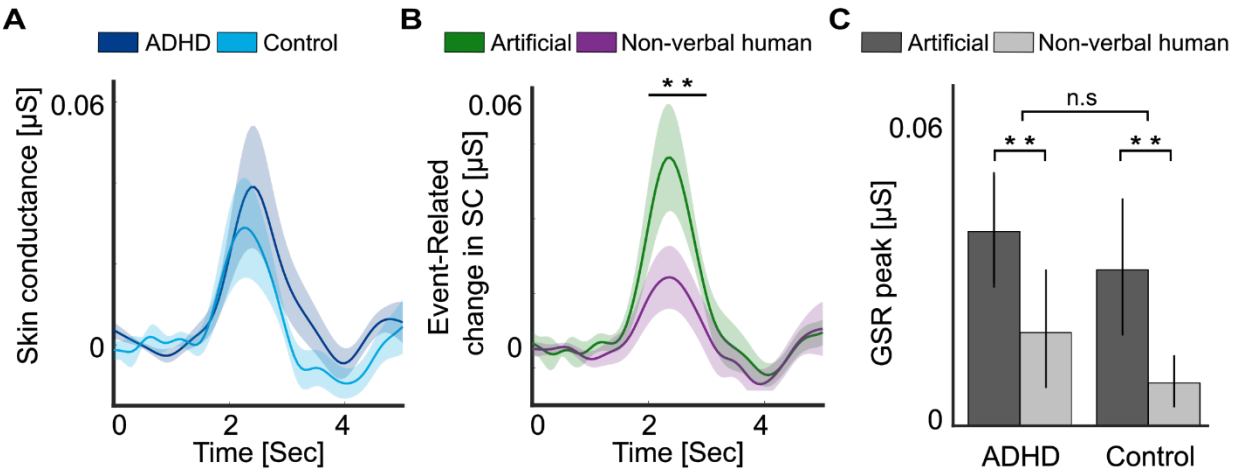


Figure 7. Event-Related changes in Skin Conductance (SC). A&B. Time course of event-related changes in phasic SC following the sound-events, shown separately for each group (A) and for the two sound-events types (B). Shaded areas around waveforms represent the SEM. Horizontal line represents the time-windows where significant differences were found. **C.** Average levels of event-related changes in phasic SC (peak between 2-3sec) shown for each group and for Artificial vs. Non-verbal human sound-events. Error bars represent the SEM. **p<0.01.

536

537 <u>Event-Related gaze-shifts</u>

538 We tested whether the sound-events triggered overt gaze-shifts away from the teacher, by
539 quantifying the percentage of event-sound that were followed by at least one gaze-shift, and
540 comparing this to control epochs taken from the Quiet condition. However, no significant
541 differences were found, suggesting that the sound-events did not elicit more frequent overt gaze-
542 shifts away from the teacher, relative to what might be observed when no sound-events are present
543 [t(48) = 1.34, p = 0.18; Figure 8A]. We also tested whether the likelihood of performing event-
544 related gaze-shifts was different for the two Event types (Artificial vs. non-verbal human sounds)
545 or in the two Groups (AD(H)D vs. Control) but did not find any significant main effects or
546 interactions [*Group*: F(47) = 0.18, p = 0.67, η^2 = 0.003 ; *Condition*: F(47) = 2.25, p = 0.14, η^2 =
547 0.045 ; *Interaction*: F(47) = 0.51, p = 0.47, η^2 = 0.01; Figure 8B].

548

549

550

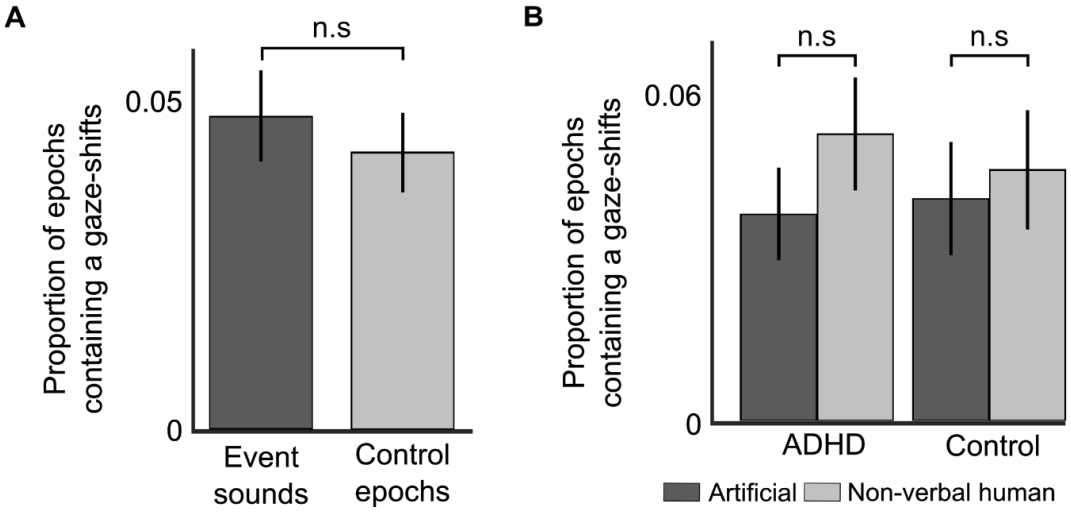


Figure 8. Event-related gaze shifts.AD(H)D **A.** Bar graph showing the number of gaze-shifts performed in 2-sec epochs following event-sounds vs. control epochs, averaged across groups and sound types. No significant differences were found, suggesting that event-sounds were not more likely to elicit overt gaze-shifts. **B.** 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.

551

552 ***Neurophysiological metrics associated with overall attention and arousal***

553 Besides analyzing neurophysiological metrics that can be directly associated with processing the
554 teacher's speech or with response to the sound-events, several additional neurophysiological
555 measures have been associated more broadly with the listeners' state of attention or arousal.
556 These including spectral properties of the EEG (and specifically power in the alpha and beta
557 range), the overall level of tonic and phasic skin-conductance, and spontaneous gaze-patterns.
558 We tested whether these metrics differed between the two groups, separately in the Quiet
559 condition and Events conditions.

560 <u>Spectral EEG features</u>

561 Spectral analysis of the EEG focused on the two frequency bands for which clear peaks were
562 detected in the periodic power-spectrum, after applying the FOOOF algorithm: Alpha (8-12Hz)
563 and Beta power (18-25Hz; Figure 9A). Power in each frequency band was assessed for each
564 participant at their personal peak within a pre-selected cluster of electrodes where each response
565 was maximal (Figure 9). We 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 either frequency band (Figure 9C) [*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)].

Global Skin Conductance Levels

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 9 D&E): 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)].

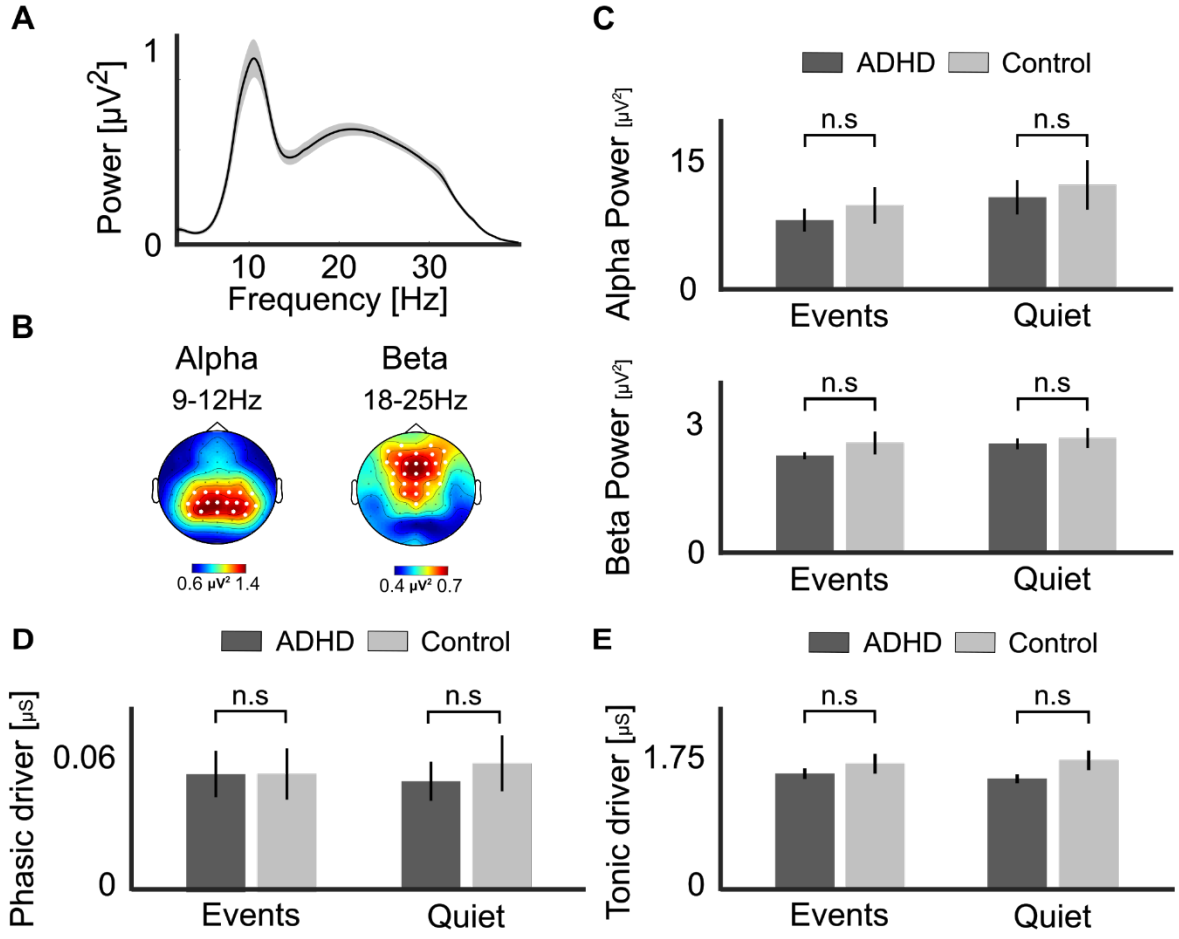


Figure 9. Spectral EEG and Skin Conductance Analysis. A. Grand-Average Power spectral density (PSD) of the Periodic portion of the EEG signal (after applying the FOOOF algorithm), shows two clear peaks corresponding to the alpha (8-12Hz) and beta (15-25Hz) bands. Shaded areas around waveforms represent 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.** Average alpha- and beta-power in Group (AD(H)D vs. Control) and condition (Quiet and Events). **D.** Phasic and **E.** tonic skin conductance levels in the same groups and conditions. No significant between-group difference was found in any comparison (n.s.). Bar graphs represent the mean values, and error bars represent the SEM.

Multivariate Analyses

The univariate analyses testing for between-group differences separately for each metric were complemented with a multivariate regression analysis, aimed at understanding the relative contribution of each neurophysiological metric to explaining between-group effects. For brevity, we focused on 10 key measures as described in the Methods section.

The pairwise Spearman's correlations between all measures are shown in Figure 10. None of the correlations were significant after correcting for multiple comparisons (fdr-correction), however three pairs did reach significance before correction (Speech-Decoding vs. ERP-P2: r = -0.33, p=0.021; TRF-N1 vs. event-related SC: r = -0.3, p=0.033; ERP-N1 vs. event-related SC: r = 0.29, p=0.04). Given these generally weak correlations, we consider these measures to be independent for the purpose of multiple regression analysis.

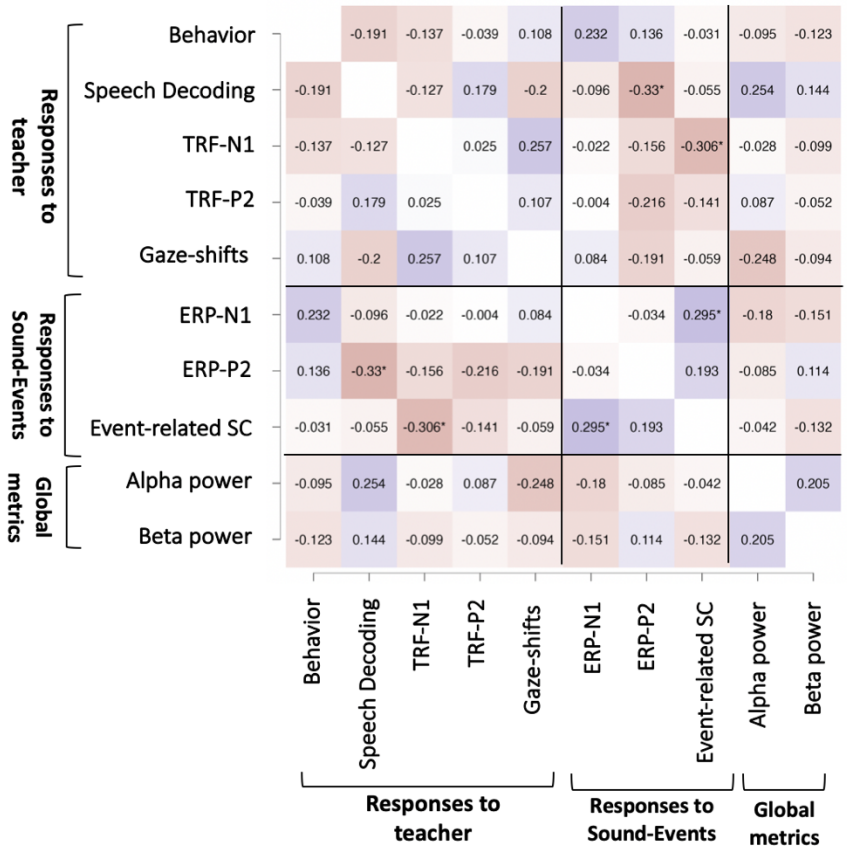


Figure 10. Correlation matrix. Heatmap of the pairwise spearman's correlation coefficient between the different neurophysiological measures included in the multivariate analysis. Red-shading indicates negative correlation values and blue-shading indicate positive values. Asterisks* indicate correlation values that pass a non-corrected threshold for statistical significance, however none survived fdr-correction for multiple comparisons.

To test the contribution of each measure for predicting whether an individual was in the ADHD or control group we performed a logistic multiple regression analysis. An 'omnibus' model containing all 10 measures achieved χ^2=15.6 with an AUC of 0.807 for predicting group-allocation. We assessed the relative contribution ('dominance') of each measure by comparing the AUC of the omnibus model to 10 models with a single measure held-out. Table 1 shows ΔAUC

contributed by each measure, indicating that the two most dominant measures were the ERP-N1 (positive contribution) and Speech Decoding (negative contribution), which were also the only two measures whose addition to the model reduced the AIC. This result is consistent with our univariate results, and indicates that these two measures contribute independently to between-group differences. Together, these two measures contributed an AUC of 0.15 to the omnibus model, which corresponds to half of the model's fit ($\Delta\chi^2$=11.8, p=0.003 relative to a model with both measures held-out).

Factor	**Dominance (ΔAUC)**	**ΔAIC**
ERP-N1*	0.054	-5.13
Speech Decoding*	0.05	-3.29
TRF-P2	0.02	+0.18
Beta power	0.01	+1.57
ERP-P2	0.005	+0.48
Event-Related SC	-0.005	+1.69
TRF-N1	0.009	+1.42
Gaze-shifts	0.007	+1.64
Alpha power	0.007	+1.6
Behavior	0.007	+1.6

Table 1: Result of the dominance analysis of a multivariate logistic regression, describing the contribution of each measure for predicting whether an individual was in the ADHD or control group.

Since ADHD is arguably not a binary category but symptoms of attentional difficulties vary on a continuum, we complemented the logistic regression analysis with a multivariate linear regression analysis, using ASRS scores as a continuous dependent measure, reflecting the severity of ADHD symptoms. An 'omnibus' model containing all 10 achieved F(10,38)=1.48 (p=0.18) with an R^2 of 0.28 for predicting ASRS scores. We assessed the relative contribution ('dominance') of each measure by comparing the R^2 in the omnibus model (containing all 10 measures) to 10 models with a single measure held-out. Table 2 shows ΔR^2 and ΔAIC contributed by each measure (smaller values are better). Interestingly, this analysis identified slightly different

621 factors than the logistic regression analysis where individuals were grouped based on having a
622 prior diagnosis of AD(H)D (Table 2). The most dominant measures, who contributed at least ΔR^2
623 >0.3 the model, were the ERP-N1 ($\beta = 0.3$), Gaze-shifts ($\beta = 0.288$), Alpha power ($\beta = 0.2$) and
624 ERP-P2 ($\beta = -0.182$). Although the omnibus model was not significant, (see above), a multivariate
625 regression model that contained only the 4 most dominant measures did significantly predicted
626 ASRS scores (R^2=0.227, p = 0.021), and collectively contributed ΔR^2=0.225 (p=0.032) to the
627 omnibus model.
628

Factor	**ΔR^2**	**ΔAIC**
ERP-N1	0.087	-3.58
Gaze-shifts	0.065	-2.23
Alpha power	0.038	-0.49
ERP-P2	0.033	-0.22
Behavior	0.028	-0.14
Speech Decoding	0.021	+0.58
TRF-N1	0.006	+1.58
Event-Related SC	0.001	+1.91
TRF-P2	0.008	+1.18
Beta power	0	+1.97

629
630 Table 2: Result of the dominance analysis of a multivariate linear regression, describing the contribution
631 of each measure for predicting individual ASRS scores.
632
633
634

Discussion

Recent years have seen a reckoning in the field of AD(H)D research, acknowledging that most current tools used for clinical assessments and research do not adequately capture the attentional challenges faced in real life (Hall et al., 2016; Barkley, 2019; Mulraney et al., 2022; Arrondo et al., 2023; Schweitzer and Zion Golumbic, 2023). We offer a new approach for advancing attention-research, by utilizing VR to simulate real-life experiences together with comprehensive measurement of different neurophysiological responses. This combination affords a more objective and well-rounded description of how individuals deal with the plethora of stimuli and task demands of their environments. Here we focused on VR classroom learning, since this is the context where many individuals, and particularly those with AD(H)D, experience difficulties in sustaining attention and avoiding distraction. Specifically, we evaluated neurophysiological measures related to focusing attention on the teacher and responses to irrelevant background sound-events alongside metrics associated more generally with levels of attention/engagement. We found that individuals diagnosed with AD(H)D exhibited significantly larger N1 response to background sound-events and somewhat reduced neural tracking of the teacher's speech. These were also the two most dominant factors contributing to a multivariate classifier that could significantly predict whether an individual was diagnosed with AD(H)D. When considering the severity of AD(H)D symptoms as a continuous variable, variance across individuals was best explained by the combined contribution of the N1 and P2 response to sound-events, amplitude of alpha-power oscillations and the frequency of gaze-shifts away from the teacher. Together, these findings emphasize the variety of attentional challenges that individuals with AD(H)D may experience in realistic scenarios and provide novel insights into the underlying neurophysiological mechanisms.

Neurophysiological responses in the VR classroom associated with symptoms of AD(H)D

The finding with the largest effect size in this study was larger N1 response to background sound-events in individuals with AD(H)D. This response showed a significant univariate between-group effect, and was the most dominant contributor to predicting group-classification (AD(H)D vs. control) as well as the severity of AD(H)D symptoms. 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). Interestingly, most studies comparing auditory ERPs in individuals with and without AD(H)D, using more traditional paradigms (e.g. oddball), typically *have not reported* differences in early-sensory components such as the N1, but rather sometimes find group differences for later ERP responses (Barry et al., 2003; Gumenyuk et al., 2005; Kaiser et al., 2020; Peisch et al., 2021). As discussed below, we attribute these discrepancies to the non-ecological nature of most traditional paradigm. The current result is in line with increased sensory sensitivity to background events, which might be especially pertinent under ecological condition. In addition, we found that neural tracking of the teacher's speech was reduced in the AD(H)D group, and this metric also emerged as a dominant predictor of AD(H)D diagnosis in multivariate between-group classification. Neural speech-tracking of continuous speech is known to be modulated by attention, and is reduced 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). We are not aware of previous studies that directly investigated neural speech tracking in AD(H)D, however these results are consistent with reduced levels of sustained attention towards the teacher in this group.

Reduced sustained attention and higher proneness to distraction are, of course, considered key characteristics of AD(H)D, however are rarely demonstrated under realistic conditions, using ecological tasks and materials. The current results demonstrate how, by coupling the ecological validity and flexibility of VR with recordings of neural activity, we can substantially advance our mechanistic understanding of attention and distractibility and obtain objective metrics of stimulus processing in realistic scenarios. Interestingly, since the magnitude of the N1 to sound-events and speech-tracking of the teacher were not correlated with each other, these results do not support the notion of an inherent 'tradeoff' between paying attention to the teacher and responding to irrelevant stimuli. Rather, they may represent two independent ways in which individuals with AD(H)D differ from their peers in the manner they respond to and process stimuli in dynamic environments.

However, adopting a binary perspective of AD(H)D can be misleading, as it does not capture the vast variability within groups, as emphasized by the RDoC framework (Morris et al., 2022). This is evident from inspecting the distribution of ASRS scores in the current sample, which shows that although the two groups are clearly separable from each other, they are far from uniform in

698 the severity of symptoms experienced. For this reason, we conducted a continuous multiple
699 regression analysis, to identify which neurophysiological measured explained variance in self-
700 reported AD(H)D symptoms. This analysis confirmed the link between the magnitude of the N1-
701 response to sound-events to AD(H)D symptoms, but also identified several additional factors that
702 contribute to explaining variance in ASRS scores across individuals, primarily: the level of alpha-
703 power and the frequency of gaze-shifts away from the teacher. This finding is consistent with the
704 hypothesized role of these metrics in attention. Higher alpha power is associated with reduced
705 levels of attention/arousal, and is consistently found to increase in conditions of boredom, tiredness
706 or prolonged time on tasks (Clarke et al., 2001; Dockree et al., 2007; Palva and Palva, 2007;
707 Wöstmann et al., 2017; Boudewyn and Carter, 2018; Jin et al., 2019; Haro et al., 2022). Several
708 studies have also found higher baseline alpha power in individuals with AD(H)D vs. controls
709 (Barry et al., 2003; Johnstone et al., 2013; Bozhilova et al., 2022; Michelini et al., 2022), although
710 results are not always consistent (Loo and Makeig, 2012; Johnstone et al., 2013). Similarly, gaze-
711 shifts are often used as a proxy for distraction, particularly under naturalistic conditions ('t Hart et
712 al., 2009; Foulsham et al., 2011; Risko and Kingstone, 2011; Risko et al., 2012; Hoppe et al.,
713 2018), and recent VR and real-life eye-tracking studies have found increased frequency and
714 prolonged duration of gaze-shifts towards irrelevant locations in individuals with AD(H)D (Braga
715 et al., 2016; Türkan et al., 2016; Vakil et al., 2019; Mauriello et al., 2022; Stokes et al., 2022). The
716 current results extend these findings to more ecological contexts and indicate that individuals who
717 exhibited higher alpha-power, performed more frequent gaze-shifts and had larger N1 response to
718 background sound-events in our VR classroom were more likely to report experiencing attentional
719 difficulties in real-life.

720 It is important to note that although the current results had significant explanatory power of
721 variance in AD(H)D symptoms, they are far from exhaustive with substantial variance still
722 unexplained by these neurophysiological measures. In this regard it is important to bear in mind
723 that the measures of AD(H)D themselves (diagnosis and ASRS scores) are also not entirely
724 reliable, given the diversity diagnostic practice (Hall et al., 2016; Barkley, 2019; Mulraney et al.,
725 2022; Arrondo et al., 2023; Schweitzer and Zion Golumbic, 2023) and the subjective nature of
726 self-reports (which are prone to bias (e.g., recency, salience, and generalization effects (Rabiner et
727 al., 2010; Toplak et al., 2013; Brevik et al., 2020)). What is clear, though, is that no single
728 neurophysiological measure alone is sufficient for explaining differences between the individuals

– whether through the lens of clinical diagnosis or through report of symptoms. These data emphasize the complex nature of attentional functioning, especially under realistic conditions, and that fact that "attentional challenges" may manifest in a variety of different ways, indexed though different neurophysiological responses. Moving forward, we believe that by measuring multiple aspects of neural and physiological responses, and studying these in diverse and ecologically valid conditions, will ultimately lead to distilling specific neurophysiological patterns that can reliably generalize across contexts.

What about the P300?

Notably, the current results diverge from previous studies of distractibility in AD(H)D in several ways. In traditional ERP studies, distractibility by irrelevant stimuli is often associated with a family of late-latency positive neural responses (the P300 family), and specifically the P3a ERP component, which 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). Given this vast literature, we had assumed that background sound-events in our VR classroom would also elicit this response. Moreover, past studies have shown reduced P3a responses in individuals with AD(H)D (Barry et al., 2003; Gumenyuk et al., 2005; Kaiser et al., 2020; Peisch et al., 2021; Kwasa et al., 2023), leading us to expect a similar pattern here. However, here the early N1 response to sound-events was larger in the ADHD group but no differences were found for any of later responses. Moreover, visual inspection of the ERPs to Artificial and non-verbal Human sound-events illustrates that not all ecological sounds elicit the same time-course of responses in mid- to late time windows.

When comparing the current results from the VR classroom with more traditional ERP studies of distractibility, we must consider the vast differences in context and task requirements and the 'functional status' of background sound-events in these studies. 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

760 participants' endogenous attention, rather than truly 'background' stimuli (Makov et al., 2023).
761 This differs substantially from the role of the novel sounds-events in the VR Classroom task used
762 here, where sound-events are clearly in the background– both spatially and semantically - and
763 occur in parallel to an ongoing speech stimulus (the lecture) that participants are primarily paying
764 attention to. In addition, the perceptual and cognitive demands of the VR Classroom – where
765 participants are instructed to listen and understand the speech as they would in a real-life classroom
766 - differ substantially from the highly vigilant demands of speeded-response tasks. These important
767 differences in task and stimulus characteristic ultimately lead to a different treatment of novel
768 sound-events by the brain, emphasizing the importance of considering context when comparing
769 results across experimental designs (Makov et al., 2023; Mandal et al., 2024).
770
771 ***Null results – insufficient sensitivity or true lack of differences?***
772 Besides the lack of a reliable P300-like effect here, several additional measures that were
773 evaluated here did not show any group-wise effects and did not contribute to explaining ASRS
774 scores, despite previous literature suggesting that they might. Although null-results are difficult to
775 interpret fully, we find it worthwhile to note these here and briefly discuss potential reasons for
776 the lack of group-related effects. Perhaps the most surprising null-effect was the lack of differences
777 in behavioral performance. Participants in both groups achieved similar behavioral performance
778 in the Quiet and Events condition, suggesting that the presence of occasional sound-events
779 ultimately did not ultimately impair their ability to understand the content of the mini-lectures (at
780 least at the level probed by our comprehension questions). These results should be interpreted
781 bearing in mind the details of the current task-design – we deliberately chose a task that would
782 mimic the level of information processing required in real-life learning contexts (as opposed to
783 speeded-response tasks used in traditional attention research that encourage failures of
784 performance by creating perceptually 'extreme' conditions (Hall et al., 2016; Barkley, 2019;
785 Arrondo et al., 2023). Moreover, the presence of occasional sound-events – while noticeable, was
786 not design to be extremely adverse. Under these conditions, individuals with and without AD(H)D
787 were able to maintain good performance despite the occasional acoustic disturbances (Mandal et
788 al., 2024), indicating that neurophysiological indication of somewhat reduced attention do not
789 necessarily translate to poorer behavioral outcomes. It is of course possible that larger behavioral

effects would have emerged if task-demands or the severity of acoustic disturbances were more challenging.

Besides neural evoked-responses, background sound-events also elicited a transient increase in SC that is associated with an 'orienting reflex', reflecting transient changes in arousal/alertness (Frith and Allen, 1983; Zimmer and Richter, 2023). This response is considered obligatory and automatic, and does not require allocation of attention as it can be observed during sleep and other unconscious states (albeit the magnitude of these responses can vary as a function of attention). Although we had hypothesized that this response might be heightened in individuals with AD(H)D, as has been reported in some studies (Sergeant, 2000; Bellato et al., 2020) (similar to the effect on the N1 response), the current results do not support this. Indeed, there has been much debate regarding the relationship between the SC orienting reflex and early neural responses, with conflicting results reported across studies (Lim et al., 1996; MacDonald and Barry, 2020). Additional research is likely required to better characterize the relationship between neural and physiological responses to surprising events in the environment.

Last, another interesting null-result was the lack of overt-gaze shifts following sound-events. Gaze-shifts are often considered a proxy for distractibility and capture of attention (Parkhurst et al., 2002; Geisler and Cormack, 2011; Nissens et al., 2016), particularly in individuals with AD(H)D. Accordingly, we had expected to find that sound-event elicited overt shifts towards the spatial location from which they were emitted. However, this was not the case, and individuals (in both groups) did not perform more gaze-shifts following sound-events relative to their baseline tendencies. A similar result was recently reported by our group in a VR café scenario, where that hearing semantically-salient words in background speech (e.g., your name or semantic violations) did elicited neural and SC responses, but was not accompanied by gaze-shifts (Brown et al., 2023). These results might indicate that peripheral auditory stimuli are, generally, less effective than visual stimuli at capturing overt attention (Turoman and Vergauwe, 2023; Mandal et al., 2024), or that in ecological audiovisual settings individuals are less prone to move their eyes to search for the source of auditory disturbances, despite noticing them. These possibilities will be further investigated in follow-up VR studies, comparing the effects of irrelevant auditory and visual events.

820 ***General considerations***

821 Before concluding, we note several important caveats, pertaining to the specific sample
822 tested here. First, as noted, classification of the AD(H)D group was based on their prior clinical
823 diagnosis, which likely varied across participants in terms of diagnosis-approach and age of
824 diagnosis. Although the two groups were distinguishable in their self-reported ASRS scores, a
825 more reliable assessment of the severity of current AD(H)D symptoms might have been achieved
826 if we had conducted a full clinical assessment of all participants, as per the acceptable clinical
827 guidelines (American Psychiatric Association, 2022). Second, this study was conducted on adults,
828 whereas the majority of AD(H)D research – and particularly all VR classroom studies - have
829 focused on children(Pollak et al., 2009; Neguț et al., 2017; Schweitzer and Rizzo, 2022; Seesjärvi
830 et al., 2022; Stokes et al., 2022). AD(H)D is primarily defined as a childhood condition and
831 although it often persists into adulthood, with profound implications on professional and
832 academics outcomes, how the nature of the deficit changes with age is still unknown (Barkley et
833 al., 2008). Moreover, the adult cohort tested here included primarily university students, who do
834 not represent all adults but rather are a sub-group with relatively high cognitive and learning
835 abilities, who are accustomed to sitting in lectures for long periods of time (Henrich et al., 2010).
836 Hence, it is possible that the lack of group-differences in some metrics in the current dataset (and
837 particularly in their behavioral outcomes) is because these individuals have learned to deal with
838 distraction and have developed appropriate strategies, which allow them to thrive in academic
839 settings. This caveat highlights the importance of considering sampling techniques in group-
840 comparison studies, which sometimes limits their interpretation to more circumscribed sub-
841 samples (Rad et al., 2018). At the same time, even if the current findings pertain only to this highly-
842 functioning academic subgroup of individuals with AD(H)D, they imply an impressive ability to
843 adapt and cope with the naturally occurring distractions of realistic classrooms. Specifically, they
844 suggest that withstanding distraction in real-life contexts may be improved through training and
845 good habit-forming, even in individuals with AD(H)D. We hope to follow up on this research in
846 the future, to deepen our understanding of the diversity within the AD(H)D and typical population
847 and to better characterize attention and distractibility across different types of real-life scenarios.
848

Conclusions

Paying attention in class requires both investing cognitive resources in processing the lesson itself (e.g., the teacher's speech), and trying to minimize intrusions from background events. Here we found that both of these operations are less effective in individuals who are diagnosed with AD(H)D or who report more severe symptoms of difficulties in everyday attention. Specifically, these individuals exhibit heightened sensory response to irrelevant sounds, reduced speech tracking of the teacher, a tendency towards more frequent gaze-shifts away from the teacher, and somewhat increased alpha-power. Interestingly, since these metrics were not correlated with each other, they likely represent different ways through which inattention or distraction may manifest, rather than a general 'biomarker' of inattention. Counter to some prevalent notions, these data suggest that AD(H)D may not be characterized by a specific "neural marker" indicating attentional deficits, particularly when tested under ecological circumstances (Barkley, 2019; Faraone et al., 2021; Sonuga-Barke et al., 2023). Rather, they resonate with recent calls to revisit the parameters used for defining 'attention-disorders' and the need to adopt more nuanced and dimensional approach, as well as designing more ecologically-valid means of testing, in order to better characterize and understand the complex nature of real-life attentional challenges(Marcus and Barry, 2011; Heidbreder, 2015; Morris et al., 2022; Elahi et al., 2023).

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).

Data Sharing Statement

The EEG and GSR data are publicly available on the Open Science Framework at: https://osf.io/svjqg.

References

Anderson NC, Bischof WF, Kingstone A (2023) Eye tracking in virtual reality. Curr Top Behav Neurosci 65:73–100.

American Psychiatric Association (2022) Diagnostic 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

921 rating scale (wurs) symptom checklists in discriminating between adults with and without
922 adhd. Brain Behav 10:e01605.
923 Brown A, Pinto D, Burgart K, Zvilichovsky Y, Zion-Golumbic E (2023) Neurophysiological
924 evidence for semantic processing of irrelevant speech and own-name detection in a virtual
925 café. Journal of Neuroscience 43:5045–5056.
926 Chennu S, Noreika V, Gueorguiev D, Blenkmann A, Kochen S, Ibáñez A, Owen AM,
927 Bekinschtein TA (2013) Expectation and attention in hierarchical auditory prediction.
928 Journal of Neuroscience 33:11194–11205.
929 Clarke AR, Barry RJ, McCarthy R, Selikowitz M (2001) Excess beta activity in children with
930 attention-deficit/hyperactivity disorder: an atypical electrophysiological group. Psychiatry
931 Res 103:205–218.
932 Coleman B, Marion S, Rizzo A, Turnbull J, Nolty A (2019) Virtual reality assessment of
933 classroom – related attention: an ecologically relevant approach to evaluating the
934 effectiveness of working memory training. Front Psychol 10:1851.
935 Crosse MJ, Zuk NJ, Di Liberto GM, Nidiffer AR, Molholm S, Lalor EC (2021) Linear modeling
936 of neurophysiological responses to speech and other continuous stimuli: methodological
937 considerations for applied research. Front Neurosci 15:705621.
938 Ding N, Simon JZ (2012a) Emergence of neural encoding of auditory objects while listening to
939 competing speakers. Proceedings of the National Academy of Sciences 109:11854–11859.
940 Ding N, Simon JZ (2012b) Neural coding of continuous speech in auditory cortex during
941 monaural and dichotic listening. J Neurophysiol 107:78–89.
942 Dockree PM, Kelly SP, Foxe JJ, Reilly RB, Robertson IH (2007) Optimal sustained attention is
943 linked to the spectral content of background eeg activity: greater ongoing tonic alpha
944 (approximately 10 hz) power supports successful phasic goal activation. Eur J Neurosci
945 25:900–907.
946 Donoghue T, Haller M, Peterson EJ, Varma P, Sebastian P, Gao R, Noto T, Lara AH, Wallis JD,
947 Knight RT, Shestyuk A, Voytek B (2020) Parameterizing neural power spectra into periodic
948 and aperiodic components. Nat Neurosci 23:1655–1665.
949 Elahi H, Iosif AM, Mukherjee P, Hinshaw SP, Schweitzer JB (2023) Using hot and cool
950 measures to phenotype and predict functional outcomes across dimensions of adhd and
951 typical development in adolescents. Res Child Adolesc Psychopathol:1–15.
952 Escera C, Alho K, Winkler I, Näätänen R (1998) Neural mechanisms of involuntary attention to
953 acoustic novelty and change. J Cogn Neurosci 10:590–604.
954 Escera C, Yago E, Corral M-J, Corbera S, Nuñez MI (2003) Attention capture by auditory
955 significant stimuli: semantic analysis follows attention switching. European Journal of
956 Neuroscience 18:2408–2412.
957 Esterman M, Rothlein D (2019) Models of sustained attention. Curr Opin Psychol 29:174–180.
958 Faraone S V. et al. (2021) The world federation of adhd international consensus statement: 208
959 evidence-based conclusions about the disorder. Neurosci Biobehav Rev 128:789–818.
960 Foulsham T, Walker E, Kingstone A (2011) The where, what and when of gaze allocation in the
961 lab and the natural environment. Vision Res 51:1920–1931.
962 Frith CD, Allen HA (1983) The skin conductance orienting response as an index of attention.
963 Biol Psychol 17:27–39.
964 Geisler WS, Cormack LK (2011) Models of overt attention. Oxford University Press.
965 Gloss D, Varma JK, Pringsheim T, Nuwer MR (2016) Practice advisory: the utility of eeg
966 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.
JASP Team (2024) JASP (version 0.19.0).
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.
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.
Kwasa JA, Noyce AL, Torres LM, Richardson BN, Shinn-Cunningham BG (2023) Top-down auditory attention modulates neural responses more strongly in neurotypical than adhd young adults. Brain Res 1798:148144.

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.
Levy O, Korisky A, Zvilichovsky Y, Zion Golumbic E (2024) The neurophysiological costs of learning in a noisy classroom: an ecological virtual reality study. The Journal of Cognitive Neuroscience in press.
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.
Mesik J, Wojtczak M (2023) The effects of data quantity on performance of temporal response function analyses of natural speech processing. Front Neurosci 16:963629.
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.

Türkan BN, Amado S, Ercan ES, Perçinel I (2016) Comparison of change detection performance and visual search patterns among children with/without adhd: evidence from eye movements. Res Dev Disabil 49–50:205–215.

Turoman N, Vergauwe E (2023) The effect of multisensory distraction on working memory: a role for task relevance? J Exp Psychol Learn Mem Cogn.

Vakil E, Mass M, Schiff R (2019) Eye movement performance on the stroop test in adults with adhd. J Atten Disord 23:1160–1169.

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, 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.