Looking for Consistency in an Uncertain World: Test-Retest Reliability of Neurophysiological and Behavioral Readouts in Autism

Shlomit Beker
Yeshiva University Albert Einstein College of Medicine
John J. Foxe
University of Rochester School of Medicine and Dentistry
John Venticinque
Yeshiva University Albert Einstein College of Medicine
Juliana Bates
Yeshiva University Albert Einstein College of Medicine
Elizabeth M. Ridgeway
Yeshiva University Albert Einstein College of Medicine
Roseann C. Schaaf
Jefferson University: Thomas Jefferson University
Sophie Molholm (sophie.molholm@einsteinmed.org)
Yeshiva University Albert Einstein College of Medicine

Research

Keywords: ASD, ICC, biomarkers, inter-trial variability, ERP, EEG

Posted Date: February 25th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-240084/v1

Looking for consistency in an uncertain world:
Test-retest reliability of neurophysiological and behavioral readouts in autism

Shlomit **Beker** [1,2], John J. **Foxe** [1,2,3], John **Venticinque** [4], Juliana **Bates** [1],

Elizabeth M. **Ridgeway** [1], Roseann C. **Schaaf** [5] and Sophie **Molholm** [1,2,3,6]

1) The Cognitive Neurophysiology Laboratory, Department of Pediatrics, Albert Einstein College of Medicine, Bronx, New York, USA

2) Department of Neuroscience, Albert Einstein College of Medicine, Bronx, New York, USA

3) The Cognitive Neurophysiology Laboratory, The Ernest J. Del Monte Institute for Neuroscience, Department of Neuroscience, University of Rochester School of Medicine and Dentistry, Rochester, New York, USA

4) School of Medicine, Albert Einstein College of Medicine, Bronx, New York, USA

5) Department of Occupational Therapy, Jefferson College of Health Professions Faculty, Farber Institute for Neurosciences Thomas Jefferson University Philadelphia, Pennsylvania, USA

6) Department of Psychiatry and Behavioral Sciences, Albert Einstein College of Medicine, Bronx, New York, USA

Corresponding Author:

sophie.molholm@einsteinmed.org

ABSTRACT

Background: Autism spectrum disorders (ASD) are associated with altered sensory processing and perception. Scalp recordings of electrical brain activity time-locked to sensory events (event-related potentials; ERPs) provide precise information on the time-course of related altered neural activity, and can be used to model the cortical loci of the underlying neural networks. Establishing the test-retest reliability of these sensory brain responses in ASD is critical to their use as biomarkers of neural dysfunction in this population.

Methods: EEG and behavioral data were acquired from 33 children diagnosed with ASD aged 6-9.4 years old, while they performed a child friendly task at two different time-points, separated by an average of 5.2 months. In two blocked conditions, participants responded to the occurrence of an auditory target that was either preceded or not by repeating visual stimuli. Intraclass correlation coefficients (ICC) were used to assess test-retest reliability of measures of sensory (auditory and visual) ERPs and performance, for the two experimental conditions. To assess the degree of reliability of the variability of responses within individuals, this analysis was performed on the variance of the measurements, in addition to their means. In total, this yielded a total of 25 measures for which ICCs were calculated.

Results: The data yielded significant good-to-excellent ICCs for 15 of the 25 measurements. These spanned across behavioral and ERPs data, experimental conditions, and mean as well as variance measures. Measures of the visual evoked responses accounted for a disproportionately large number of the significant ICCs.

Conclusions: This analysis reveals that sensory ERPs and related behavior can be highly reliable across multiple measurement time-points in ASD. The data further suggest that the inter-trial and inter-participant variability reported in the ASD literature likely often represents replicable individual participant neural processing differences. The stability of these neuronal readouts supports their use as biomarkers in clinical and translational studies on ASD. Given the minimum interval between test/retest sessions across our cohort, we also conclude that for the tested age-range of ~6 to 9.4 years these reliability measures are valid for at least a 3-month interval.

Key words: ASD, ICC, biomarkers, inter-trial variability, ERP, EEG

BACKGROUND

Autism Spectrum Disorder (ASD) is defined by social-communication deficits and restricted and repetitive patterns of behavior, and is often accompanied by sensory, motor, perceptual and cognitive atypicalities. Although well defined by clinical diagnostic criteria and assessed professionally through interviews and clinical observation, ASD is highly heterogeneous, with wide ranging presentation and a variety of etiologies and developmental trajectories (1-3). As a neurodevelopmental condition, direct measures of brain activity provide for greater understanding of the underlying neuropathology and how this impacts information processing. If robust, replicable and reliable neurophysiological measures of processing differences in ASD can be developed, these might then have utility in the stratification of individuals at early stages of the condition, to optimize targeted interventions, and as biomarkers for assaying treatment efficacy.

Scalp recordings of electrophysiological brain responses (Electroencephalogram: EEG) provide a non-invasive readout of network level neural processing with millisecond temporal resolution. EEG time-locked to stimulus presentation or to behavioral responses, referred to as event-related potentials (ERPs), is used to characterize the time-course of information processing (4, 5), and can also be used to model the cortical loci of the underlying neural networks (6, 7). EEG/ERPs are thus well-suited to the characterization of when and where cortical information processing might be altered in ASD, and have the potential to provide sensitive assays of treatments that are expected to act on processes with a clear neural signature. Additionally, since EEG/ERPs directly index neural function, they are likely more sensitive to initial treatment effects, given that they can measure site-of-action effects in real-time. This feature is particularly meaningful for clinical trials, which tend to be of relatively short-duration, and would benefit from more sensitive and immediate outcome measures. In contrast, more typical clinical and behavioral assays might be expected to show

somewhat delayed treatment-related changes, since neural changes due to intervention would only give rise to changes in behavioral outcomes after sufficient time has passed.

There is an accumulation of support for altered sensory-perceptual processing in ASD, with evidence for differential processing across all the major sensory modalities, including audition (8-12), vision (13-16), somatosensation (17, 18) and multisensory integration systems (19). However, it bears mentioning that these differences, when present, can often be subtle, and that there has tended to be a high degree of inconsistency across the literature [see (20)]. Nonetheless, a promising development is that variance in sensory ERPs has been related to the severity of the clinical phenotype (12, 21), going to their utility as potential biomarkers. A similarly promising development is work showing that both auditory and visual sensory responses can be modulated by training, signifying potential sensitivity to treatment effects (22, 23). However, sensory ERPs have not yet been submitted to standard assessment of test-retest reliability in ASD, which is surely a minimal requirement in assessing their potential as sensitive biomarkers. Indeed, this seems particularly germane given often inconsistent findings across studies, and suggestions by some research groups of increased inter-trail variability of the sensory-evoked response in ASD (2, 24), but see (25, 26).

In the quest for reliable biomarkers to index brain function in ASD, we sought here to measure the test-retest reliability of auditory and visual evoked potentials and related task performance. High-density EEG recordings and behavioral responses were recorded from children with ASD while they engaged in a simple speeded reaction time task in response to visually cued and non-cued auditory stimuli. Intraclass correlation coefficients (ICC) were calculated to assess the reliability of the sensory evoked responses and behavioral data recorded across two identical experimental sessions that were temporally separated by an average of about 5 months (27-29).

Participants: Data from 33 children diagnosed with ASD ranging from 6.1 to 9.4 years of age were included for this analysis (see Table 1 for participant characteristics). These came from a larger dataset (N=94) collected in the context of a clinical trial on the efficacy of different behavioral interventions, and included a subset of the participants from whom we recorded EEG and behavioral data from two sessions, which we refer to as test (pre intervention) and retest (post intervention).

While we had full datasets from test and retest sessions in 40 participants, 7 (17.5%) were excluded from the current analysis because of insufficient data due to artifact contamination in one or both of the recording sessions. See Table 2 for reasons for exclusion and for comparison of the demographics and characteristics of the included versus excluded participants; see also Discussion section.

Table 1: Mean±SD and range of characteristics and cognitive scores of all participants.

Sex	Age	Time between tests (Months)	Handedness	IQ* (Full scale)	IQ (Verbal)	IQ (Non Verbal)	ADOS severity
29(M) 4 (F)	7.54 ± 1 [6.1 – 9.4]	Average: 5.2 ± 2 [2.9 – 10.4]	Right/Left: 20/13	92.9 ± 16 [58 – 131]	90.4 ± 20 [56 – 130]	95.5 ± 16 [64 – 128]	7.78 ± 1.5 [5 – 10]

Table 2: Reasons for exclusion (A), and statistical comparison of clinical and other characteristics of included and excluded participants (B)

A. Reasons for exclusion

	Withdrew from the st after 1st session	Covid-19 related issues	EEG attempted but unsuccessful	Noisy EEG data	Still in intervention
N=61	8 (13.1%)	13 (21.3%)	31 (50.8%)	7 (11.5%)	2 (3.3%)

B. Age, Sex and cognitive scores

	Age	Sex	IQ (Full scale)	IQ (Verbal)	IQ (Non Verbal)	ADOS severity
Included in analysis (N=33)	7.54 ± 1 [6.1 – 9.4]	29(M) 4 (F)	92.9 ± 16 [58 – 131]	90.4 ± 20 [56 – 130]	95.5 ± 16 [64 – 128]	7.78 ± 1.5 [5 – 10]
Excluded from analysis (N=61)	7.46 ± 1 [6 – 9.5]	50 (M) 11 (F)	83.8 ± 19 [50 - 124]	75.9 ± 22 [45 - 128]	92.5 ± 19 [56 - 143]	8.1 ± 1.7 [3 – 10]
Difference (two sample T-test)	t-stat = 0.1 df = 92 p = 0.87	χ^2stat=0.56 p = 0.45	t-stat = 2.6 df = 92 p = 0.01	t-stat = 3 df = 92 p = 0.003	t-stat = 0.8 df = 92 p = 0.42	t-stat = 0.4 df = 92 p = 0.63

The time between test and retest was 5.2 ± 2 months (Min: 2.9; max: 10.4). Participants were recruited without regard to sex, race or ethnicity. IQ quotients for performance (PIQ), verbal (VIQ), and full-scale (FSIQ) intelligence were assessed in all of the participants using the Wechsler Abbreviated Scales of Intelligence (WASI; (30)). To be considered for the study, participants had to meet diagnostic criteria for ASD on the basis of the Autism Diagnosis Observation Schedule (ADOS-2) (31), childhood history, and clinical impression of a licensed clinician with extensive experience in the evaluation and diagnosis of children with ASD. The Repetitive Behavior Scale-Revised (RBS-R) (32) questionnaire was collected to obtain continuous measures of ASD characteristics related to insistence on sameness such as ritualistic/sameness behavior, stereotypic behavior, and restricted interests. Participants received modest recompense for their participation (a total of $250 for participation in the treatment study). Exclusionary criteria included epilepsy or premature birth (<35 weeks). While the majority of participants had performance IQs>80, a subset (N=6) had lower scores (ranging from 64 to 79, with a mean of 73 and standard deviation of 5.3). All participants passed a screen for normal or corrected-to-normal vision and normal hearing on the day of testing. Parents and/or guardians of all participants provided written informed consent. All procedures were approved by the Institutional Review Board of the Albert Einstein College of Medicine.

Stimuli and Task: Visual cue stimuli were presented on a 25" ViewSonic screen (refresh rate: 60 Hz, pixel resolution: 1280x1024x32) using Presentation© software (Version 20.0, Neurobehavioral Systems, Inc., Berkeley, CA, www.neurobs.com), running on a Dell computer. Visual stimuli consisted of a still image of a cartoon dog face located at the center of the screen, subtending ~4.4° of visual angle. Auditory stimuli were delivered at an intensity of 75 dB SPL via a single, centrally-located loudspeaker (JBL Duet Speaker System, Harman Multimedia) (See Figure 1 for paradigm schematic and the corresponding grand average ERP responses over the full trial epoch at occipital channels).

The task was designed to test the hypothesis that children with ASD do not use temporally predictive information in a typical way (33, 34). The task included two conditions: For the *Cue* condition, participants were presented with a sequence of 4 visual isochronous stimuli for a duration of 80ms each presented at a Stimulus Onset Asynchrony (SOA) of 650ms, followed by an 80ms auditory stimulus, presented 650ms after the onset of the last visual cue. For the *No-Cue* control condition, the auditory stimulus was not preceded by a sequence of visual cue stimuli. Both conditions included 15% catch trials on which the auditory target was not presented. Each target appeared 2600ms after the beginning of the trial, during which participants were focused on the screen. In all other respects, the paradigm, including the timing of the stimuli, was identical between the Cue and No-Cue conditions. Cue and No-Cue conditions were presented in blocks, with 25 trials per block, and a total of 20 blocks (10 Cue; 10 No-Cue). The order of blocks within the experiment for a given participant was randomly generated prior to each experimental session. Each block lasted 3.5 minutes. Participants were encouraged to take short breaks between blocks as needed. The entire experimental session lasted around 3 hours, and, in addition to data acquisition, included cap application, frequent short breaks, lunch, and cap removal. Participants were seated at a fixed distance of 65 cm from the screen and responded with their preferred hand. In all trials, they were instructed to press a button on a response pad (Logitech Wingman Precision Gamepad) as soon as they heard the auditory tone. Responses occurring between 150-1500ms after the auditory target stimulus were considered valid, and positive feedback was provided via presentation of a cartoon dog image and an uplifting sound. If the response was outside this time window, a running dog cartoon with a sad sound was presented to indicate that the response was too fast, and a sitting dog image with the sad sound was presented to indicate that the response was too slow. Frequent breaks were given as needed to ensure maximal task concentration. Here we focus on the sensory evoked and behavioral responses to evaluate their reliability between two data recording sessions separated by

a minimum of 10 weeks (2.5 months). Analyses to test the hypothesis that children with ASD do not use temporally predictive information in a typical way are presented in another report in which we compare ASD data to an age matched typically developing control group (34).

Data processing: Data were processed and analyzed using custom MATLAB scripts (MATLAB r2017a, MathWorks, Natick, MA), and the FieldTrip toolbox (35). A minimum number of 50 EEG trials per condition per analysis was set as a criterion for a participant to be included in the analysis, however, most participants had more than 100 trials in each condition (test: mean ± standard deviation (SD): 208±86; retest: 209±95). Due to occasionally long reaction times in some of the participants, only responses given within 1000ms after stimulus presentation were considered as valid for further analysis of the behavioral data.

Measurements that were used in the ICC calculations were calculated as follows:

1. **Behavior:**

   Reaction times and sensitivity, indexed by D-Prime (d') (36, 37), scores were calculated from the behavioral data. Hits were defined as responses that occurred between 150 ms to 1500 ms following the auditory tone. False alarms were defined as a response to a catch trial – i.e. pushing the button even though no auditory target stimulus occurred. D' was calculated for each participant: $d' = Z\big(p(Hit)\big) - Z(p(False\ Alarm))$. ICC was calculated for RT means and SDs, and for D', for both Cue and No-Cue conditions.

2. **EEG Data:**

   Continuous EEG data were down-sampled to 256 Hz, band-pass filtered between 0.1 and 55 Hz using Butterworth Infinite Impulse Response (IIR) windowing with filter order of 5, and then epoched as specified below. Epochs were demeaned to normalize for DC shifts, and baseline-corrected using the 100 ms time window prior to stimulus onset.

i. **Visual Evoked Response (VEP)**: To derive the VEP, epochs of 200ms before and 850ms after visual stimulus presentation were generated and baselined to the 100ms pre stimulus onset, and then averaged across trials separately for each participant and recording session. Data were referenced to a midline frontal channel (AFz) to optimize visualization and measurement of the VEP over occipital scalp at channels O1, O2, and Oz. Comparison of the VEP between the sessions was calculated on the voltage at the peak of the P1, N1, and P2 components for each participant within time windows of 80ms, 60ms, and 100ms, centered at 100ms (visual P1), 180ms (visual N1) and 350ms (visual P2), respectively. Both means and SDs per participant were used to assess ICC for mean and inter-trial variability metrics of the VEP, respectively, for a total of 6 measures across the three visual components.

ii. **Auditory Evoked Response (AEP)**: To derive the AEP, epochs of 300ms before and 850ms after auditory stimulus presentation were generated and baselined to the 100ms pre stimulus onset, and then averaged across trials separately for each participant and recording session. Data were referenced to a channel near the left mastoid (TP7) to optimize visualization and measurement of the AEP over fronto-central scalp. Statistical analyses were performed on data from fronto-central channels (FC1, FC2, FCz), at the peak of each participant's auditory P1, N1 and P2 component (time windows of 50ms, 70ms and 80ms centered at 50ms (auditory P1), 100ms (auditory N1) and 200ms (auditory P2), respectively. Both means and SDs per participant were used to assess ICC for mean and inter-trial variability metrics of the AEP. This yielded a total of 12 measures since these were generated for both Cue and No-Cue conditions for each of three components.

iii. **Phase measurement**: We were also interested in whether mean phase angle at the Cue condition stimulation frequency of 1.5 Hz was consistent for participants across testing

session, since oscillatory phase alignment has been shown under rhythmic stimulation conditions (38-42). This was measured at the onsets of visual stimuli, at 1.5 Hz, and yielded a single measure for ICC analysis. Phase was calculated for each participant and session on a trial-by-trial basis, on the complex number achieved through Morlet-based wavelet convolution of the signal. To encompass the full sequence of stimuli comprising a trial, EEG data were epoched at 3000ms before and 500ms after the auditory event. Trials were baselined to the 100ms before the onset of the first visual stimulus in the sequence, and referenced to a frontal channel (AFz). Analysis of phase was performed on epochs that were low-pass filtered at 55Hz, and high pass filtered at 0.1Hz. Following previous studies showing posterior entrainment to rhythmically presented visual inputs (38, 40), our areas of interest for phase alignment were focused on parieto-occipital channels (PO3, POz, PO4).

Test-retest analysis

Our analyses focused on assessing the consistency of behavioral and electrophysiological responses across two recording sessions. To do this, we performed Inter-Class Correlation Coefficient (ICC) analyses using a one-way mixed effect model with absolute agreement and multiple observations (27, 29, 43, 44), according to the formula:

$$ICC(1,k) = \frac{MS_R - MS_w}{MS_R}$$

MS_R = mean square for rows (variance between participants); MS_W = mean square for residual sources of variance; *k* = number of raters (measurements).

Separate ICCs were calculated for test-retest pairs for each of the 25 measurements. ICC was computed with the Intraclass Correlation Coefficient package:

https://www.mathworks.com/matlabcentral/fileexchange/22099-intraclass-correlation-coefficient-icc, MATLAB Central File Exchange (Arash Salarian, 2020). To correct for multiple comparisons, Bonferroni correction for multiple comparisons (45) was applied to the ICC R values.

For purely descriptive purposes, Pearson linear correlation coefficients were calculated for the pairs of observations (test/retest) for each of the behavioral and ERP parameters. While such correlations do not account for absolute agreement between the values across sessions as does ICC, they allow for visualization of the relationship between test and retest measures. Pearson correlation was computed as: $\rho = \frac{cov(X,Y)}{\sigma X \cdot \sigma Y}$. Cov = covariance of test and retest; σX and σY is the SD of test and retest, respectively. To control for False Discovery Rate (FDR), Bonferroni correction for multiple comparisons (45) was applied on all p-values of all correlations.

Testing for association between test-retest similarity and participant cognitive variables: First, a test-retest similarity index (SI) was calculated for each individual, indicating the degree of similarity between the test and retest across all measurements. SI was calculated on all test-retest pairs as following:

$$Y = Z_{score}(X_1 \dots X_n), \quad SI_n = 1 - var(\sum_{k=i}^{m} {Y_n}/{m})$$

$X =$ Measurement (ERP or behavior), *n*=number of participants, *m*=number of measurements (ERP and behavior).

To measure for possible associations between the SI and participant cognitive variables, Pearson correlation coefficients were calculated between SI, PIQ, VIQ, RBSR and ADOS, in the form of a correlation matrix. Results were then corrected for multiple comparisons (45).

Finally, to test for the possibility that test-retest reliability found for the participants was linked to the time that had passed between the sessions, which varied quite widely between 2.9 and 10.4

months, we measured the correlation between the participants' similarity index and the time interval between the test and the retest.

RESULTS

The auditory and visual sensory evoked responses at test and retest are illustrated in grand average VEP and AEP waveforms and topographic maps in Figure 2, and the individual participant VEP, AEP and behavioral responses in Figure 3. A striking similarity between the group mean responses can be observed in Figure 2, whereas at the individual participant level some variance is apparent (Figure 3). ICC analyses were performed to formally assess the consistency of responses at the individual participant level between test and retest.

Intraclass Correlation Coefficient (ICC) results

25 measures of the EEG and behavioral data (see methods) were submitted to ICC analysis. ICC values are presented in Figure 4. ICC R values, P values and lower and upper bounds of the 95% confidence interval, calculated separately for each measurement, are presented in Table 3.

For ICC analysis, the higher the R-value is, the stronger the agreement between the two sessions. Per convention, values below 0.50 are generally considered to have a poor level of reliability, values from 0.50 to 0.75 to be of moderate reliability, values from 0.75 to 0.90 to have good reliability, and, when higher than 0.90 they are considered to have excellent reliability (46). Accordingly, the following measurements had good reliability: RT Cue, RT No-Cue, RT ITV No-Cue, VEP P1, VEP P1 ITV, VEP N1, VEP N1 ITV, VEP P2 ITV, AEP N1 Cue, and AEP P2 Cue. After correcting for multiple comparison with Bonferroni correction, four measures that were found as significant prior to the correction were no longer significant: AEP P1 No-Cue ITV; RT Cue ITV; d' Cue; d' No-Cue. Using more liberal criteria to categorize ICC values as proposed by Fleiss (47), in which $R < 0.40$ as poor, $0.40 < R < 0.75$ as fair to good, and $R > 0.75$ as excellent, the top 10 measures in Table 3 would be considered excellent, rather than good. Pearson correlations yielded significant correlations for all

measurements but phase angle (uncorrected p<0.05; see Figure 5). d' cue, d' No-Cue, RT cue, RT No-Cue, RT ITV No-Cue, VEP P1, VEP P1 ITV, VEP N1, VEP N1 ITV, VEP P2, VEP P2 ITV, AEP N1, AEP N1 ITV, AEP P2, and AEP N1 No-Cue remained significant following correction for multiple comparisons. Pearson correlations for the test-retest pairs are presented in Figure 5.

Table 3: ICC results

Measure	R value (ICC)	P value	UB	LB
ICC > 0.75				
VEP N1	*0.8615*	*1.2359x10^-7*	*0.9321*	*0.7185*
VEP N1 ITV	*0.856*	*1.9969x10^-7*	*0.9294*	*0.7072*
RT NC ITV	*0.8282*	*1.1297x10^-6*	*0.9149*	*0.6548*
RT NC	*0.8201*	*1.9527x10^-6*	*0.9108*	*0.6385*
VEP P2 ITV	*0.8148*	*3.8236x10^-6*	*0.9093*	*0.6236*
RT Cue	*0.811*	*3.4601x10^-6*	*0.9063*	*0.6203*
VEP P1	*0.7946*	*2.4445x10^-5*	*0.8993*	*0.5825*
VEP P1 ITV	*0.7944*	*1.2216x10^-5*	*0.8992*	*0.582*
AEP N1 Cue	*0.7926*	*1.2074x10^-5*	*0.8984*	*0.5785*
AEP P2 Cue	*0.7518*	*8.7397x10-^5*	*0.8784*	*0.4955*
0.5 < ICC < 0.75				
AEP N1 NC	*0.7462*	*1.0912x10^-4*	*0.8757*	*0.4842*
VEP P2	*0.7203*	*2.8088x10^-4*	*0.8629*	*0.4314*
AEP N1 Cue ITV	*0.7181*	*3.0223x10^-4*	*0.8619*	*0.427*
AEP P2 NC ITV	*0.6657*	*0.0014*	*0.8362*	*0.3204*
AEP P2 Cue ITV	*0.6616*	*0.0016*	*0.8342*	*0.3121*
D' No-Cue	0.6136	0.004	0.8084	0.2234
AEP P1 NC ITV	0.6114	0.0048	0.8096	0.2102
RT Cue ITV	0.6082	0.0045	0.8058	0.2125
D' cue	0.595	0.0058	0.7992	0.186
ICC < 0.5				
AEP N1 NC ITV	0.4451	0.0513	0.7281	-0.128
AEP P1 Cue	0.4387	0.0547	0.725	-0.1408
AEP P1 NC	0.3302	0.1324	0.6718	-0.3615
AEP P2 NC	0.2632	0.1973	0.639	-0.4976
Phase	0.2152	0.253	0.6201	-0.6142
AEP P1 Cue ITV	0.2088	0.2567	0.6123	-0.6081

Table 3. R-values, P values, and upper bounds (UB) and lower bounds (LB) of the 95% confidence interval for test-retest, for each of the measurements, ranked from the highest ICC value to the lowest. In italics: significant measurements after correction for multiple comparisons.

Similarity Index (SI) and clinical measures

SI was generated for each participant (see methods) and tested for correlation with ADOS severity scores, PIQ, VIQ and RBSR in a correlation matrix (Figure 6). None of these correlations survived Bonferroni correction (45). Finally, there was no correlation between SI and time between the sessions (Rho=0.006; p=0.97)

DISCUSSION

In autism research, several factors bring into question the possibility that brain measurements can serve as reliable markers of neurocognitive function. Basic findings on sensory processing from recordings of electrophysiological brain activity often differ across laboratories; and there is some evidence of higher inter-participant (48-50) and inter-trial ((51-53), but see (25, 26)) variability within such recordings compared to control groups. This raises the possibility that such measurements may simply be too noisy to serve as reliable readouts of brain function in ASD. Alternatively, differences in findings between laboratories may result from factors that do not have direct implications for the reliability of the scalp recorded electrical brain response, such as differences in stimuli, task, EEG recording setup and analysis pipeline, ascertainment bias and clinical cohort. What is more, inter-participant variability may reflect a feature of the heterogeneity of the condition rather than random noise. Surprisingly few studies to date have sought to test the stability of these responses when participants, recording equipment, analytic approach, and stimulation parameters are held constant, which is particularly critical to establish if a biomarker is to be used as an outcome measure in a clinical trial, or as a reliable indicator of neural and neurocognitive dysfunction (54, 55). Only two previous studies, as far as we are aware, examined the reliability of such measures across two recording sessions Levin and colleagues (56) collected 5 minutes of resting state EEG from children with and without ASD at two intervals separated by ~6 days, and

found good reliability of the center frequency and amplitude of the largest alpha-band peak. Cremone-Caira and colleagues (57) found moderate to good consistency of the executive function related frontal-N2 response elicited during go/nogo and flanker tasks in children with ASD across two time points separated by ~3 months.

Here we add to this emerging literature with the finding that in children with ASD, auditory and visual ERPs, as well as reaction-times collected in an accompanying target detection task, show *good* to *excellent* test-retest reliability. We found statistically significant test-retest reliability, as measured by the Intraclass Correlation Coefficient, for a full 15 of the 25 electrophysiological and behavioral measurements submitted to analysis, with significant ICC values ranging from 0.65 to 0.86 (representing good to excellent ICC values (46, 47). Interestingly, these high ICC values were found not only for mean responses but also for the inter-trial variance of these responses. Significance was found across data category (ERP and behavior), sensory domain (VEP and AEP), ERP component (P1, N1 and P2), experimental condition (Cue and No-Cue) and response metric (mean and SD). Among the 10 measurements for which significant ICCs were not found, 6 were from the AEP (representing 50% of the AEP derived measures). In notable contrast, ICC was significant for all measurements of the VEP. Of the remaining four values that did not achieve significance, 3 were behavioral (RT ITV for the cued condition and D' for cued and non-cued conditions) and the remaining was for mean phase angle. From this we conclude that in the current setup, VEP and mean RT were most reliably consistent across test retest measurements. Given the minimum interval between test/retest sessions across our cohort, we also conclude that for the tested age-range of ~6 to 9.4 years these reliability measures are valid for at least a 3-month interval.

Atypicalities in sensory-evoked neural responses and behavioral performance have been widely reported in ASD, including altered responses to visual (2, 49, 58), auditory (59), and somatosensory

stimuli (60-62). Moreover, in some studies, higher inter-participant (48-50) and within-participant inter-trial (51-53) variability of brain responses to sensory stimuli has been shown. This is in line with higher inter-trial behavioral variability that was observed for individuals with ASD, measuring reaction times to executive function (63) and tactile judgment tasks (64), as well as rhythmic tapping tasks (65). The higher variability between trials and between individuals with ASD has, in turn, been interpreted in the context of neuronal processing being "noisy" or "unreliable" (e.g., (49, 50, 52, 66-68), but see (69, 70) for reports of lower noise in ASD, and (25, 26) for evidence of typical levels of noise in ASD). According to this view, high levels of endogenous neural noise in ASD render neural signals unreliable (53, 71). Arguing against a pure noise account, here we see a stable pattern of both mean activity and inter-trial variability over time. The current data suggest that such variance likely represents replicable neural processing differences at the individual participant level in the clinical group, rather than noise. Hypo- and/or hypersensitivity of synaptic activity, for example, could lead to a higher than typical range of neuronal responses to a given stimulus (72). A possible result would be an increased range of neural activity across large-scale neural networks, that is nevertheless stable over time (25). At the same time, given the consistency of individual responses within our clinical group, the inter-participant variability that has been observed (48, 50) is likely to reflect that ASD has a variety of etiologies and developmental routes (1), that in turn lead to heterogeneous neural and behavioral phenotypes.

A number of notable recent reviews have focused on the promise of EEG based biomarkers of IDDs, and discussed the requirements and challenges therein (54, 73-75). Biomarkers have the potential to serve many purposes including assessment of risk, diagnosis, disease progression, intervention response, and mechanism of disease. Validity and reliability of the potential biomarker are critical to establish. Here, we find good-to-excellent reliability of sensory evoked responses to simple

auditory and visual stimuli using an active paradigm suitable for children. Since auditory and visual sensory ERPs have been shown to differ in ASD, the additional finding that they can be reliably measured and show stability within individuals over time opens the door to their further development as biomarkers. Next steps will be to establish if these measures are equally reliable in the absence of a task and how they are affected by state (e.g., drowsy versus alert), to determine if they can be applied in more severely affected individuals (76-78). Given the simplicity of the paradigm and stimuli, such biomarkers could also be suitable for translational studies in non-human models of ASD (see discussion by Modi & Sahin, 2017).

We should note that full datasets were collected for fewer than half of the potential cohort. Consideration of the reasons for this, and the implications for EEG biomarker use in clinical studies, is worthwhile. The parent study, a clinical trial, required a minimum of ~40 lab visits. During these visits clinical assessments were performed, collection of primary outcome measures was made at three time-points, and therapy sessions occurred. Due to the already significant demands of the parent study, EEG recordings were not prioritized since they did not provide a primary outcome measure. In this context about half of the participants that completed the parent study did not yield full EEG datasets (N=31): 31% did not perform the task correctly or at all and so EEG data collection was terminated, 29% would not wear the cap, 16% refused to continue the EEG experiment partway into data collection, 13% did not sit still enough to acquire good EEG data and so data collection was terminated, and for <1% either no attempt at EEG data collection was made or hair style prevented adequate cap application. Participant characteristics for included and excluded participants are presented in Table 2. Most notably, verbal IQ and full scale IQ were significantly higher for the included group. Otherwise, participant demographics and characteristics appeared to be highly similar.

This brings to light possible challenges for EEG-data collection in clinical trials in pediatric populations with neurodevelopmental disorders, especially when using a paradigm in which participants perform an active task. Use of a passive auditory or somatosensory paradigm while watching a movie with the sound off, an approach that we often take with lower functioning individuals, would have obviated issues of task compliance, and may also have reduced boredom and hyperactivity. However, the impetus is on us, as researchers with the goal of developing EEG biomarkers that can be used as outcome measures in clinical trials, to develop approaches for pediatric clinical populations that allow EEG data collection in a wider range of circumstances. Low montage EEG recordings (e.g., (79, 80)), using wireless technology, and embedding of stimuli in movies or highly engaging video games or stories (81) are just some adaptations that may increase participant compliance. These are not yet commonly used, if at all. The validity and reliability of EEG measurements under such conditions are not known, and will have to be established each time significant methodological changes are introduced.

We note that when EEG data collection is primary to the study that the participant has been recruited for and therefore is prioritized, we typically have high levels of compliance of at least 85%. In our EEG studies in high functioning clinical populations, we achieve at least this rate of compliance even when we use paradigms that involve relatively complex tasks. What is more, we have similar compliance rates in our EEG studies in lower functioning populations such as Rett Syndrome (76-78) and Batton Disease, where we use passive paradigms that do not require task performance, and in individuals with severe neuropsychiatric conditions (82, 83).

ICC is a strong metric of the reliability of a response for a given group, but it does not provide individual scores that can be used to assess how test-retest similarity may vary as a function of another variable such as the time interval between measures. We therefore generated a composite measure for each individual, the Similarity Index (SI), which is simply the mean of the variance between test/retest values across all of the z-scored ERP and behavioral measures. The SI may be considered a composite measure of the stability within an individual of the neuronal/behavioral readouts, evoked by a given task. The more similar the test-retest readouts of a process are, the more stable and less variable the neuronal activity that underlies this process. The SI was used to test if the differences in the length of the interval between test and retest systematically varied with the reliability of the responses. It is important to note that differences in intervals arose due to uncontrolled factors such as appointments being rescheduled. In this context, there was no evidence for such a relationship, although we are reluctant to draw strong conclusions from this, since the study was not designed to test this hypothesis. Indeed, one might expect such a relationship, given that the brain is still immature and neuronal plasticity relatively high at these ages. We additionally tested for possible covariance of SI scores with participant traits, as represented by cognitive/clinical variables, but found no significant correlation between SI and RBSR, IQ or autism severity scores. Future work will be required to establish the validity of such a composite SI.

Study limitations

A potential limitation of the current analysis is that the data were collected in the context of a treatment study. Of the 33 participants included in these analyses, two thirds (n=22) were in active treatment groups (applied behavioral analysis (ABA), N=10; sensory integration therapy (SIT), N=12), and one-third in a *treatment as usual* control group (N=11). Importantly however, with regard to

hypotheses for the parent study, there was no expectation that the treatments would influence the basic auditory and visual sensory responses or reaction times that we focused on here and for which we found good consistency.

Another limitation of these data is that EEG was not acquired for all participants, for reasons including hyperactive behavior that prevented them from being able to sit for EEG recordings, inability to do the task, and refusal to keep the cap on. As described earlier, this highlights possible challenges for research-grade EEG data collection for clinical trials (see earlier discussion, and Table 2).

Lastly, while our study finds strong consistency of neuronal and behavioral measurements in children with ASD, it does not include similar data from an age-matched typically developing (TD) control group, and thus we cannot draw conclusions regarding whether reliability differs from a healthy control group. However, this does not detract from evidence for remarkably good consistency of the responses between two recording sessions in children with ASD and all its implications, which is a critical feature for a treatment biomarker.

Declarations

Author Contributions: SM and JJF conceptualized the study. JB oversaw and performed clinical and cognitive assessments. SB, JJF and SM consulted closely with each other on data analysis and statistical testing of data. SB wrote the scripts for the analyses and JV and SB performed all analyses (signal processing and statistical). SB made the illustrations and wrote the first substantial draft of the manuscript, in consultation with SM and JF. SB, JJF and SM, JB, EMR, and RCS all contributed to editing of the manuscript. All authors approve this final version for publication.

Acknowledgements: We extend our heartfelt gratitude to the families that contributed their time and effort to participate in this research. The authors thank Catherine Halprin and Sophia Zhou who recruited the majority of the participants, and Douwe Horsthuis and Alaina Berruti who collected the majority of the EEG data.

Funding: This work was supported in part by an RO1 from the NICHD (HD082814) and through the Rose F. Kennedy Intellectual and Developmental Disabilities Research Center (RFK-IDDRC), which is funded through a center grant from the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD U54 HD090260). Work on ASD at the University of Rochester (UR) collaboration site is supported in part through the UR Intellectual and Developmental Disabilities Research Center (UR-IDDRC), which is funded by a center grant from the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD P50 HD103536).

Data Sharing: The authors will make the full de-identified dataset with appropriate notation and any related analysis code available in a public repository (Figshare) and include digital object identifiers within the final text of the paper, so that any interested party can access them.

Conflict of interest statement: All authors confirm no competing financial interests.

Abbreviations: ASD (Autism Spectrum Disorder); EEG (Electroencephalogram); ERP (evoked response potentials); ICC (Intraclass correlation coefficients); RT (Reaction Time); SD (Standard Deviation); df (degrees of freedom); VEP (visual evoked response); AEP (auditory evoked response); ITV (inter-trial variability); NC (No-Cue)

REFERENCES:

1. Masi A, DeMayo MM, Glozier N, Guastella AJ. An Overview of Autism Spectrum Disorder, Heterogeneity and Treatment Options. Neurosci Bull. 2017;33(2):183-93.

2. Milne E, Gomez R, Giannadou A, Jones M. Atypical EEG in autism spectrum disorder: Comparing a dimensional and a categorical approach. J Abnorm Psychol. 2019;128(5):442-52.

3. Fein D, Barton M, Eigsti IM, Kelley E, Naigles L, Schultz RT, et al. Optimal outcome in individuals with a history of autism. J Child Psychol Psychiatry. 2013;54(2):195-205.

4. Picton tW, Bentin s, Berg p, Donchin e, Hillyard sA, Johnson j, r., et al. Guidelines for using human event-related potentials to study cognition: Recording standards and publication criteria. Psychophysiology. 2000;37:127-52.

5. Foxe JJ, Simpson GV. Flow of activation from V1 to frontal cortex in humans. A framework for defining "early" visual processing. Exp Brain Res. 2002;142(1):139-50.

6. Scherg M, Berg P, Nakasato N, Beniczky S. Taking the EEG Back Into the Brain: The Power of Multiple Discrete Sources. Front Neurol. 2019;10:855.

7. Scherg M, Berg P. Use of Prior Knowledge in Brain Electromagnetic Source Analysis. Brain Topography. 1991;4.

8. Bruneau N, Bonnet-Brilhault F, Gomot M, Adrien J-L, Barthélémy C. Cortical auditory processing and communication in children with autism: electrophysiological/behavioral relations. International Journal of Psychophysiology. 2003;51(1):17-25.

9. Jansson-Verkasaloa E, Ceponiene R, Kielinene M, Suominen K, Jantti V, Linnae S, et al. Deficient auditory processing in children with Asperger Syndrome, as indexed by event-related potentials. Neuroscience Letters. 2003;338:197-200.

10. Orekhova EV, Stroganova TA, Prokofiev AO, Nygren G, Gillberg C, Elam M. The right hemisphere fails to respond to temporal novelty in autism: evidence from an ERP study. Clin Neurophysiol. 2009;120(3):520-9.

11. Stroganova TA, Kozunov VV, Posikera IN, Galuta IA, Gratchev VV, Orekhova EV. Abnormal pre-attentive arousal in young children with autism spectrum disorder contributes to their atypical auditory behavior: an ERP study. PLoS One. 2013;8(7):e69100.

12. Brandwein AB, Foxe JJ, Butler JS, Frey HP, Bates JC, Shulman LH, et al. Neurophysiological indices of atypical auditory processing and multisensory integration are associated with symptom severity in autism. J Autism Dev Disord. 2015;45(1):230-44.

13. Frey HP, Molholm S, Lalor EC, Russo NN, Foxe JJ. Atypical cortical representation of peripheral visual space in children with an autism spectrum disorder. Eur J Neurosci. 2013;38(1):2125-38.

14. McPartland J, Dawson G, Webb SJ, Panagiotides H, Carver LJ. Event-related brain potentials reveal anomalies in temporal processing of faces in autism spectrum disorder. J Child Psychol Psychiatry. 2004;45(7):1235-45.

15. Sysoeva OV, Galuta IA, Davletshina MS, Orekhova EV, Stroganova TA. Abnormal Size-Dependent Modulation of Motion Perception in Children with Autism Spectrum Disorder (ASD). Front Neurosci. 2017;11:164.

16. Fiebelkorn IC, Snyder AC, Mercier MR, Butler JS, Molholm S, Foxe JJ. Cortical cross-frequency coupling predicts perceptual outcomes. Neuroimage. 2013;69:126-37.

17. Khan S, Michmizos K, Tommerdahl M, Ganesan S, Kitzbichler MG, Zetino M, et al. Somatosensory cortex functional connectivity abnormalities in autism show opposite trends, depending on direction and spatial scale. Brain. 2015;138(Pt 5):1394-409.

18. Kemner C, Verbaten MN, Cuperus JM, Camfferman G, Van Engeland H. Visual and somatosensory event-related brain potentials in autistic children and three different control groups. Electroencephalography and clinical Neurophysiology. 1994;92:225-37.

19. Brandwein AB, Foxe JJ, Butler JS, Russo NN, Altschuler TS, Gomes H, et al. The development of multisensory integration in high-functioning autism: high-density electrical mapping and psychophysical measures reveal impairments in the processing of audiovisual inputs. Cereb Cortex. 2013;23(6):1329-41.

20. Williams ZJ, Abdelmessih PG, Key AP, Woynaroski TG. Cortical Auditory Processing of Simple Stimuli Is Altered in Autism: A Meta-analysis of Auditory Evoked Responses. Biol Psychiatry Cogn Neurosci Neuroimaging. 2020.

21. Roberts TP, Khan SY, Rey M, Monroe JF, Cannon K, Blaskey L, et al. MEG detection of delayed auditory evoked responses in autism spectrum disorders: towards an imaging biomarker for autism. Autism Res. 2010;3(1):8-18.

22. Dawson G, Jones EJ, Merkle K, Venema K, Lowy R, Faja S, et al. Early behavioral intervention is associated with normalized brain activity in young children with autism. J Am Acad Child Adolesc Psychiatry. 2012;51(11):1150-9.

23. Anderson S, White-Schwoch T, Parbery-Clark A, Kraus N. Reversal of age-related neural timing delays with training. Proc Natl Acad Sci U S A. 2013;110(11):4357-62.

24. Haigh SM. Variable sensory perception in autism. Eur J Neurosci. 2018;47(6):602-9.

25. Butler JS, Molholm S, Andrade GN, Foxe JJ. An Examination of the Neural Unreliability Thesis of Autism. Cereb Cortex. 2017;27(1):185-200.

26. Coskun MA, Varghese L, Reddoch S, Castillo EM, Pearson DA, Loveland KA, et al. Increased response variability in autistic brains? Neuroreport. 2009;20(17):1543-8.

27. Shrout Patrick E. FJL. Intraclass Correlations: Uses in Assessing Rater Reliability. Psychol Bull. 1979;86(2):420-8.

28. Ebel R, L. Estimation of the Reliability of Ratings. Psychometrika. 1951;16(December).

29. Bartko JJ. The Intraclass Correlation Coefficient as a Measure of Reliability. Psychological Reports. 1966;19:3-11.

30. Stano JF. Wechsler Abbreviated Scale of Intelligence. Rehabilitation Counseling Bulletin. 2004;48:56-7.

31. Lord C, Rutter M, Le Couteur A. Autism Diagnostic Interview-Revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. J Autism Dev Disord 1994;24(5):659-85.

32. Lam KS, Aman MG. The Repetitive Behavior Scale-Revised: independent validation in individuals with autism spectrum disorders. J Autism Dev Disord. 2007;37(5):855-66.

33. van Laarhoven T, Stekelenburg JJ, Vroomen J. Suppression of the auditory N1 by visual anticipatory motion is modulated by temporal and identity predictability. Psychophysiology. 2020:e13749.

34. Beker S, Foxe JJ, Molholm S. Oscillatory entrainment mechanisms and anticipatory predictive processes in Autism Spectrum Disorder (ASD). bioRxiv. 2020:2020.05.07.083154.

35. Oostenveld R, Fries P, Maris E, Schoffelen JM. FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Comput Intell Neurosci. 2011;2011:156869.

36. Simpson AJ, Fitter MJ. What is the best index of detectability? Psychological Bulletin. 1973;80:481-8.

37. Green DM, Swets JA. Signal Detection Theory and Psychophysics. . New York: Wiley 1966.

38. Gray MJ, Frey HP, Wilson TJ, Foxe JJ. Oscillatory recruitment of bilateral visual cortex during spatial attention to competing rhythmic inputs. J Neurosci. 2015;35(14):5489-503.

39. Ten Oever S, Schroeder CE, Poeppel D, van Atteveldt N, Mehta AD, Megevand P, et al. Low-Frequency Cortical Oscillations Entrain to Subthreshold Rhythmic Auditory Stimuli. J Neurosci. 2017;37(19):4903-12.

40. Wilson TJ, Foxe JJ. Cross-frequency coupling of alpha oscillatory power to the entrainment rhythm of a spatially attended input stream. Cogn Neurosci. 2020;11(1-2):71-91.

41. Fiebelkorn IC, Foxe JJ, Butler JS, Mercier MR, Snyder AC, Molholm S. Ready, set, reset: stimulus-locked periodicity in behavioral performance demonstrates the consequences of cross-sensory phase reset. J Neurosci. 2011;31(27):9971-81.

42. Mercier MR, Molholm S, Fiebelkorn IC, Butler JS, Schwartz TH, Foxe JJ. Neuro-oscillatory phase alignment drives speeded multisensory response times: an electro-corticographic investigation. J Neurosci. 2015;35(22):8546-57.

43. McGraw KOWSP. Forming Inferences About Some Intraclass Correlation Coefficients. Psychological Methods. 1996;1(1):30-46.

44. Malcolm BR, Foxe JJ, Butler JS, Mowrey WB, Molholm S, De Sanctis P. Long-term test-retest reliability of event-related potential (ERP) recordings during treadmill walking using the mobile brain/body imaging (MoBI) approach. Brain Res. 2019;1716:62-9.

45. Dunn OJ. Multiple Comparisons Among Means. Journal of the American Statistical Association. 1961;56:52-64.

46. Portney LGaW, M.P. Foundations of Clinical Research: Applications to Practice. 3rd Edition Pearson Education, Inc, New Jersey. 2009.

47. Fleiss J. The Design and Analysis of Clinical Experiments. Wiley, New York. 1986.

48. Hahamy A, Behrmann M, Malach R. The idiosyncratic brain: distortion of spontaneous connectivity patterns in autism spectrum disorder. Nat Neurosci. 2015;18(2):302-9.

49. Kovarski K, Malvy J, Khanna RK, Arsene S, Batty M, Latinus M. Reduced visual evoked potential amplitude in autism spectrum disorder, a variability effect? Transl Psychiatry. 2019;9(1):341.

50. Park WJ, Schauder KB, Zhang R, Bennetto L, Tadin D. High internal noise and poor external noise filtering characterize perception in autism spectrum disorder. Sci Rep. 2017;7(1):17584.

51. Latinus M, Mofid Y, Kovarski K, Charpentier J, Batty M, Bonnet-Brilhault F. Atypical Sound Perception in ASD Explained by Inter-Trial (In)consistency in EEG. Front Psychol. 2019;10:1177.

52. Dinstein I, Heeger DJ, Lorenzi L, Minshew NJ, Malach R, Behrmann M. Unreliable evoked responses in autism. Neuron. 2012;75(6):981-91.

53. Milne E. Increased intra-participant variability in children with autistic spectrum disorders: evidence from single-trial analysis of evoked EEG. Front Psychol. 2011;2:51.

54. Ewen JB, Sweeney JA, Potter WZ. Conceptual, Regulatory and Strategic Imperatives in the Early Days of EEG-Based Biomarker Validation for Neurodevelopmental Disabilities. Front Integr Neurosci. 2019;13:45.

55. McPartland JC, Bernier RA, Jeste SS, Dawson G, Nelson CA, Chawarska K, et al. The Autism Biomarkers Consortium for Clinical Trials (ABC-CT): Scientific Context, Study Design, and Progress Toward Biomarker Qualification. Front Integr Neurosci. 2020;14:16.

56. Levin AR, Naples AJ, Scheffler AW, Webb SJ, Shic F, Sugar CA, et al. Day-to-Day Test-Retest Reliability of EEG Profiles in Children With Autism Spectrum Disorder and Typical Development. Front Integr Neurosci. 2020;14:21.

57. Cremone-Caira A, Vaidyanathan A, Hyatt D, Gilbert R, Clarkson T, Faja S. Test-retest reliability of the N2 event-related potential in school-aged children with autism spectrum disorder (ASD). Clin Neurophysiol. 2020;131(2):406-13.

58. Simmons DR, Robertson AE, McKay LS, Toal E, McAleer P, Pollick FE. Vision in autism spectrum disorders. Vision Res. 2009;49(22):2705-39.

59. Bonnel A, Mottron L., Peretz I, Trudel M, Gallun E, Bonnel A. Enhanced Pitch Sensitivity in Individuals with Autism: A Signal Detection Analysis. 2003.

60. Kern JK, Trivedi MH, Garver CR, Grannemann BD, Andrews AA, Savla JS, et al. The pattern of sensory processing abnormalities in autism. Autism. 2006;10(5):480-94.

61. <Kanner_1943.pdf>.

62. Kanner L. Autistic Disturbances of Affective Contact. Pathology. 1943.

63. Geurts HM, Grasman RP, Verte S, Oosterlaan J, Roeyers H, van Kammen SM, et al. Intra-individual variability in ADHD, autism spectrum disorders and Tourette's syndrome. Neuropsychologia. 2008;46(13):3030-41.

64. Puts NA, Wodka EL, Tommerdahl M, Mostofsky SH, Edden RA. Impaired tactile processing in children with autism spectrum disorder. J Neurophysiol. 2014;111(9):1803-11.

65. Morimoto C, Hida E, Shima K, Okamura H. Temporal Processing Instability with Millisecond Accuracy is a Cardinal Feature of Sensorimotor Impairments in Autism Spectrum

Disorder: Analysis Using the Synchronized Finger-Tapping Task. J Autism Dev Disord. 2018;48(2):351-60.

66. Weinger PM, Zemon V, Soorya L, Gordon J. Low-contrast response deficits and increased neural noise in children with autism spectrum disorder. Neuropsychologia. 2014;63:10-8.

67. Heeger DJ, Behrmann M, Dinstein I. Vision as a Beachhead. Biol Psychiatry. 2017;81(10):832-7.

68. Haigh SM, Heeger DJ, Dinstein I, Minshew N, Behrmann M. Cortical variability in the sensory-evoked response in autism. J Autism Dev Disord. 2015;45(5):1176-90.

69. Davis G, Plaisted-Grant K. Low endogenous neural noise in autism. Autism. 2015;19(3):351-62.

70. Brock J. Alternative Bayesian accounts of autistic perception: comment on Pellicano and Burr. Trends Cogn Sci. 2012;16(12):573-4; author reply 4-5.

71. Rubenstein J, Merzenich M. Model of autism: increased ratio of excitation/ inhibition in key neural systems. Genes, Brain and Behavior. 2003;2:255-67.

72. Chen Q, Deister CA, Gao X, Guo B, Lynn-Jones T, Chen N, et al. Dysfunction of cortical GABAergic neurons leads to sensory hyper-reactivity in a Shank3 mouse model of ASD. Nat Neurosci. 2020;23(4):520-32.

73. McPartland JC. Developing Clinically Practicable Biomarkers for Autism Spectrum Disorder. J Autism Dev Disord. 2017;47(9):2935-7.

74. Sahin M, Jones SR, Sweeney JA, Berry-Kravis E, Connors BW, Ewen JB, et al. Discovering translational biomarkers in neurodevelopmental disorders. Nat Rev Drug Discov. 2018.

75. Modi ME, Sahin M. Translational use of event-related potentials to assess circuit integrity in ASD. Nat Rev Neurol. 2017;13(3):160-70.

76. Sysoeva OV, Molholm S, Djukic A, Frey HP, Foxe JJ. Atypical processing of tones and phonemes in Rett Syndrome as biomarkers of disease progression. Transl Psychiatry. 2020;10(1):188.

77. Brima T, Molholm S, Molloy CJ, Sysoeva OV, Nicholas E, Djukic A, et al. Auditory sensory memory span for duration is severely curtailed in females with Rett syndrome. Transl Psychiatry. 2019;9(1):130.

78. Foxe JJ, Burke KM, Andrade GN, Djukic A, Frey HP, Molholm S. Automatic cortical representation of auditory pitch changes in Rett syndrome. J Neurodev Disord. 2016;8(1):34.

79. Fouad IA. A robust and reliable online P300-based BCI system using Emotiv EPOC + headset. J Med Eng Technol. 2021:1-19.

80. Bleichner MG, Debener S. Concealed, Unobtrusive Ear-Centered EEG Acquisition: cEEGrids for Transparent EEG. Front Hum Neurosci. 2017;11:163.

81. Isbell E, Wray AH, Neville HJ. Individual differences in neural mechanisms of selective auditory attention in preschoolers from lower socioeconomic status backgrounds: an event-related potentials study. Dev Sci. 2016;19(6):865-80.

82. Francisco AA, Foxe JJ, Horsthuis DJ, DeMaio D, Molholm S. Assessing auditory processing endophenotypes associated with Schizophrenia in individuals with 22q11.2 deletion syndrome. Transl Psychiatry. 2020;10(1):85.

83. Francisco AA, Horsthuis DJ, Popiel M, Foxe JJ, Molholm S. Atypical response inhibition and error processing in 22q11.2 Deletion Syndrome and schizophrenia: Towards neuromarkers of disease progression and risk. Neuroimage Clin. 2020;27:102351.

Figure Titles and Legends

Figure 1

Schematic of experimental paradigm

A) Top: Cue condition trial. Bottom: Cue condition grand average responses over trial epoch at occipital channels (O1,O2,Oz). B) Top: No-Cue condition trial. Bottom: grand average of evoked responses for the trial. No-Cue condition grand average responses over trial epoch at occipital channels (O1,O2,Oz).

Figure 2

ERPs: Visual Evoked Potentials (VEP) and Auditory Evoked Potentials (AEP) in the two test sessions

A) VEP (averaged over channels O1, O2, and Oz) collapsed across all visual evoked responses, in test (red) and retest (blue). B) Topography maps for the VEP P1 (1st row), N1 (2nd row), P2 (3rd row) components shown in A for (from right to left): test, retest, and the difference between them. C) AEP (averaged over channels FC1, FC2, FCz) for test and retest in the Cue conditions. D) Topography maps for the AEP P1, N1, P2 components in Cue condition, for test, retest, and the difference between them. E) AEP for test and retest in the No-Cue condition) Same as in D, for No-Cue condition.

Figure 3

Individual-level ERPs and reaction times (RT) for test and retest

A) Top, ERPs showing VEP (left) and AEP (right) for test and retest, for all participants (each colored line represent an evoked response of an individual participant). Black: Grand average for each session. Bottom, Illustration of measurement consistency for the ERP data that showed the highest ICC scores: Amplitudes of the visual N1 (left) and auditory N1 (right) at test and retest. B) Illustration of measurement consistency for the behavioral data that showed the highest ICC scores: Reaction times (RT) for the Cue (left) and No-Cue (middle) conditions, and Inter-trial Variability (ITV) of RT for No-Cue (right), at test and retest.

Figure 4

ICC values, grouped by measurement type: VEP, AEP (Cue and No-Cue) and Behavior

Figure 5

Pearson correlations for test-retest pairs

Results are shown for all behavioral (red) and evoked sensory (blue) and phase (green) ERP measures used in the study. Rho and p values for show significant correlations for all but high-order EEG measures. NC: No-Cue.

Figure 6

Correlation matrix of clinical scores and Similarity Index (SI)

Gray scale colors code for Pearson rho. Uncorrected P values are given for each correlation.

Figures

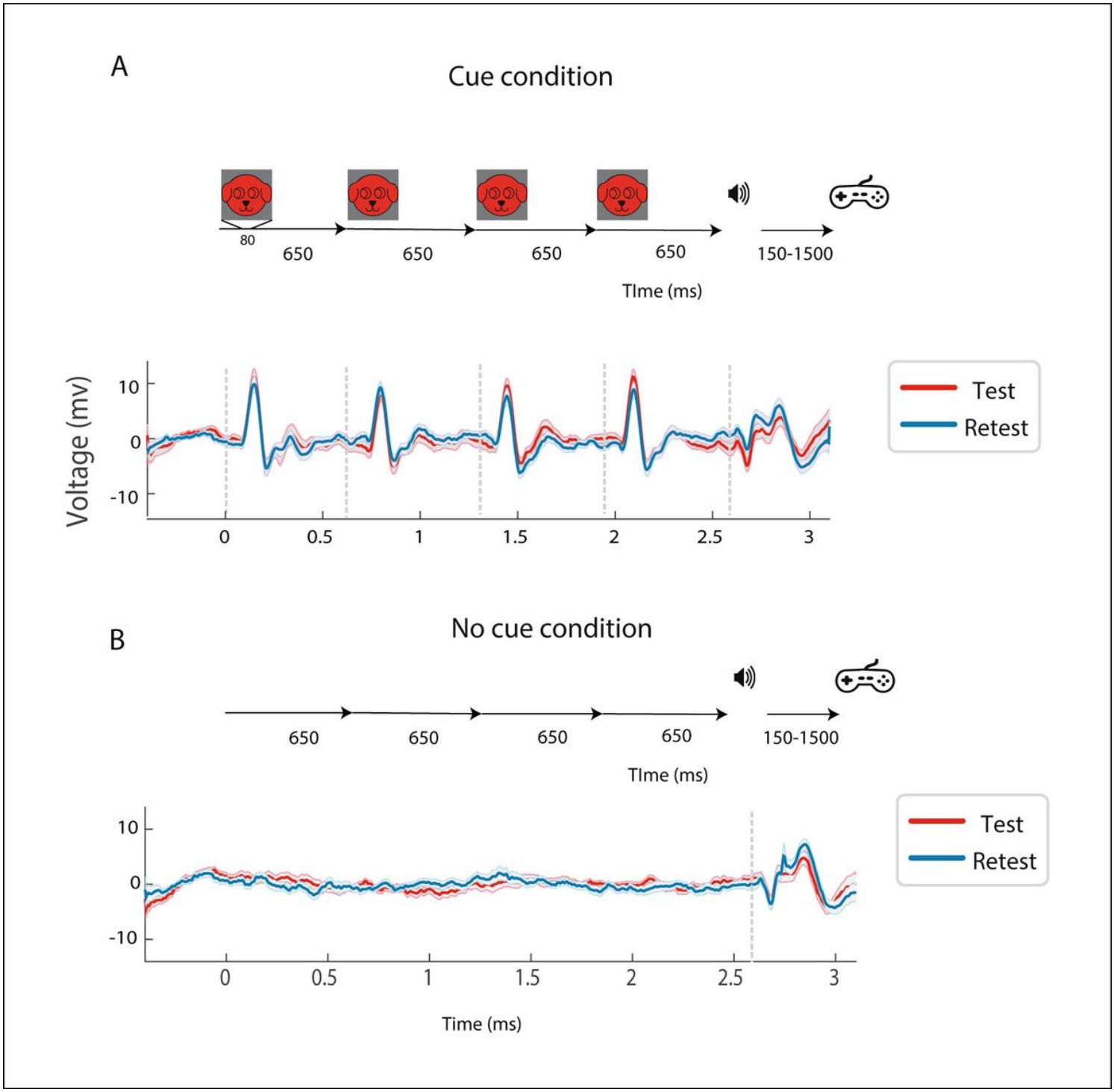

Figure 1

Schematic of experimental paradigm A) Top: Cue condition trial. Bottom: Cue condition grand average responses over trial epoch at occipital channels (O1,O2,Oz). B) Top: No-Cue condition trial. Bottom: grand average of evoked responses for the trial. No-Cue condition grand average responses over trial epoch at occipital channels (O1,O2,Oz).

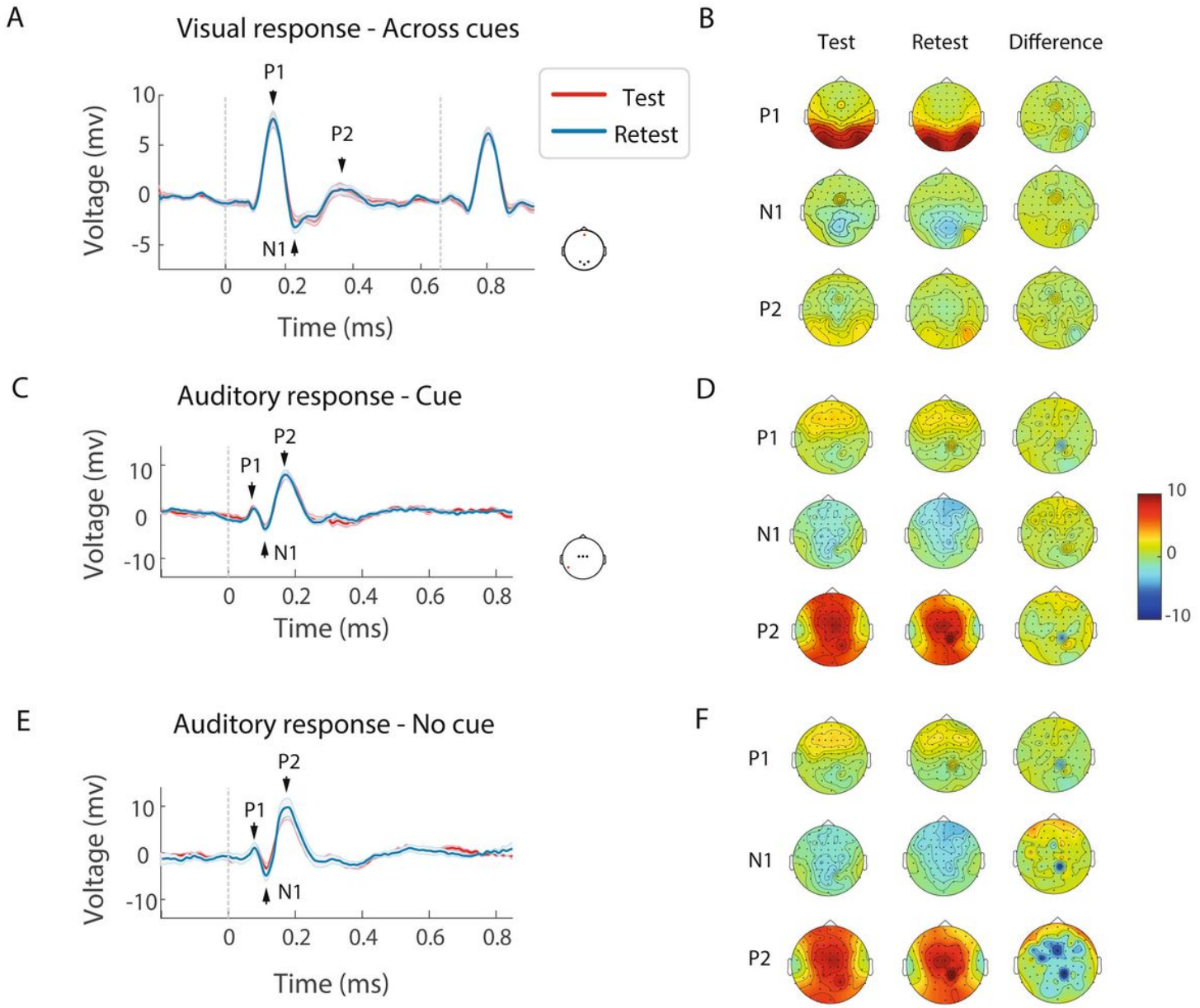

Figure 2

ERPs: Visual Evoked Potentials (VEP) and Auditory Evoked Potentials (AEP) in the two test sessions A) VEP (averaged over channels O1, O2, and Oz) collapsed across all visual evoked responses, in test (red) and retest (blue). B) Topography maps for the VEP P1 (1st row), N1 (2nd row), P2 (3rd row) components shown in A for (from right to left): test, retest, and the difference between them. C) AEP (averaged over channels FC1, FC2, FCz) for test and retest in the Cue conditions. D) Topography maps for the AEP P1, N1, P2 components in Cue condition, for test, retest, and the difference between them. E) AEP for test and retest in the No-Cue condition) Same as in D, for No-Cue condition.

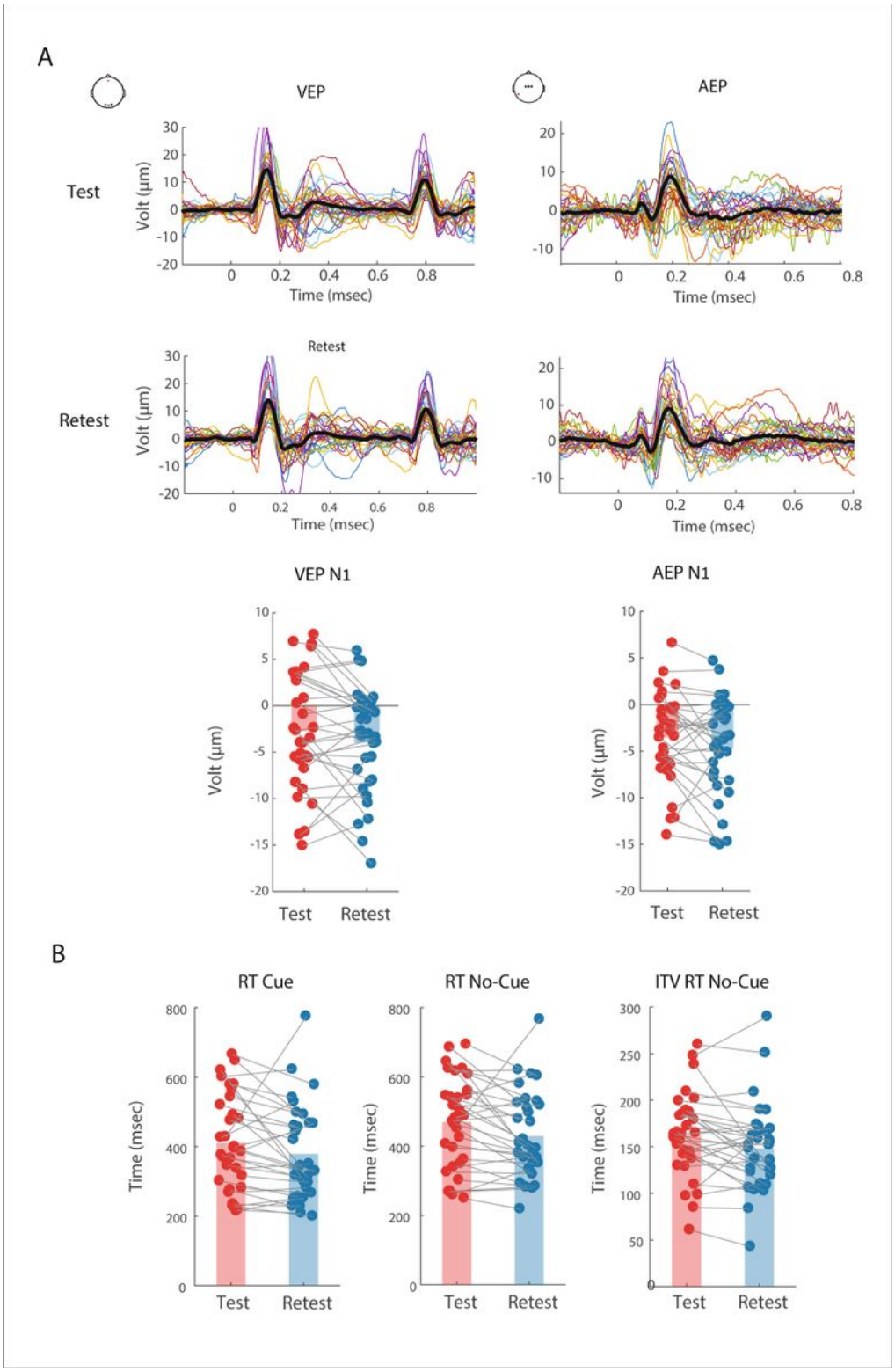

Figure 3

Individual-level ERPs and reaction times (RT) for test and retest A) Top, ERPs showing VEP (left) and AEP (right) for test and retest, for all participants (each colored line represent an evoked response of an individual participant). Black: Grand average for each session. Bottom, Illustration of measurement consistency for the ERP data that showed the highest ICC scores: Amplitudes of the visual N1 (left) and auditory N1 (right) at test and retest. B) Illustration of measurement consistency for the behavioral data

that showed the highest ICC scores: Reaction times (RT) for the Cue (left) and No-Cue (middle) conditions, and Inter-trial Variability (ITV) of RT for No-Cue (right), at test and retest.

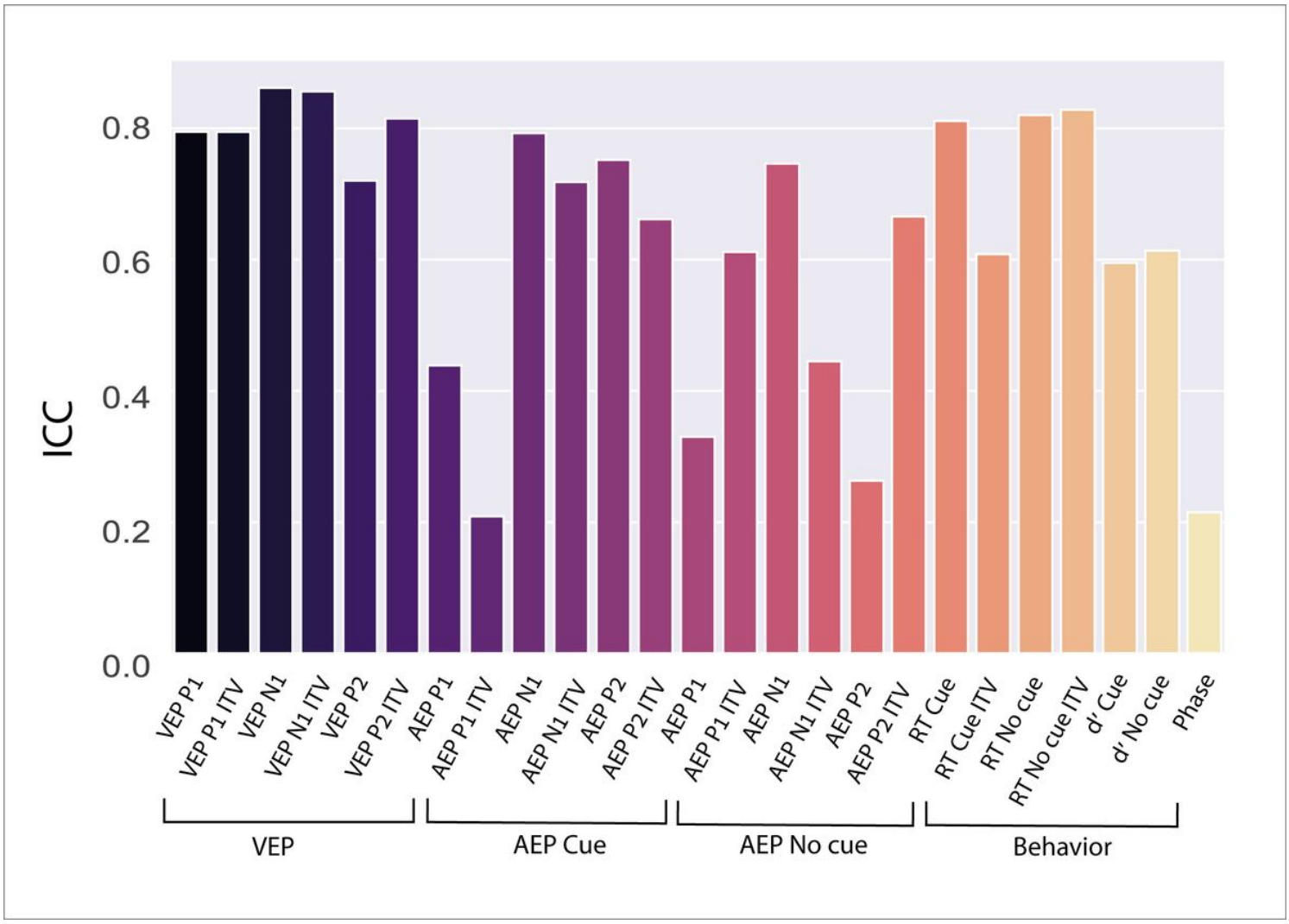

Figure 4

ICC values, grouped by measurement type: VEP, AEP (Cue and No-Cue) and Behavior

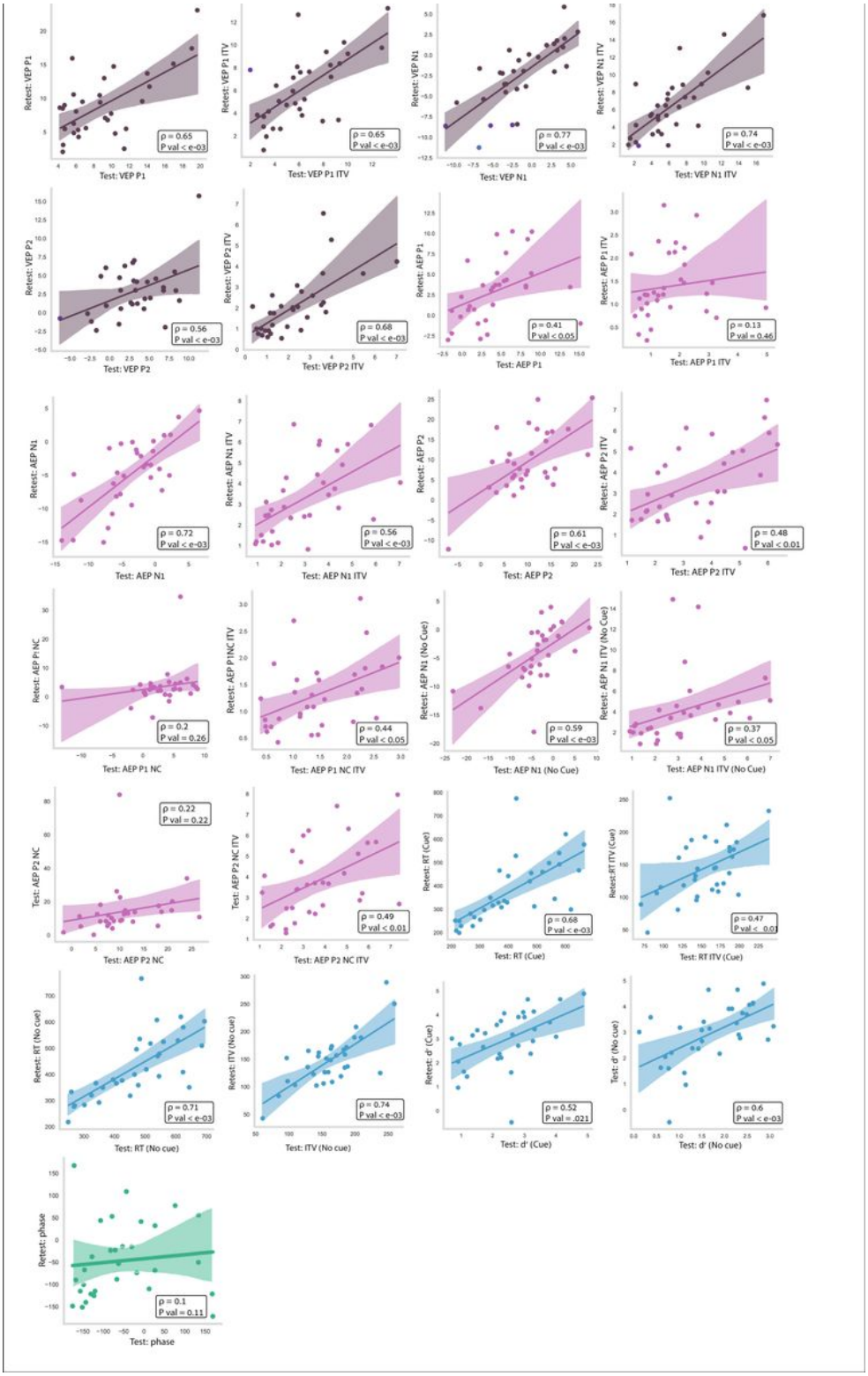

Figure 5

Pearson correlations for test-retest pairs Results are shown for all behavioral (red) and evoked sensory (blue) and phase (green) ERP measures used in the study. Rho and p values for show significant correlations for all but high-order EEG measures. NC: No-Cue.

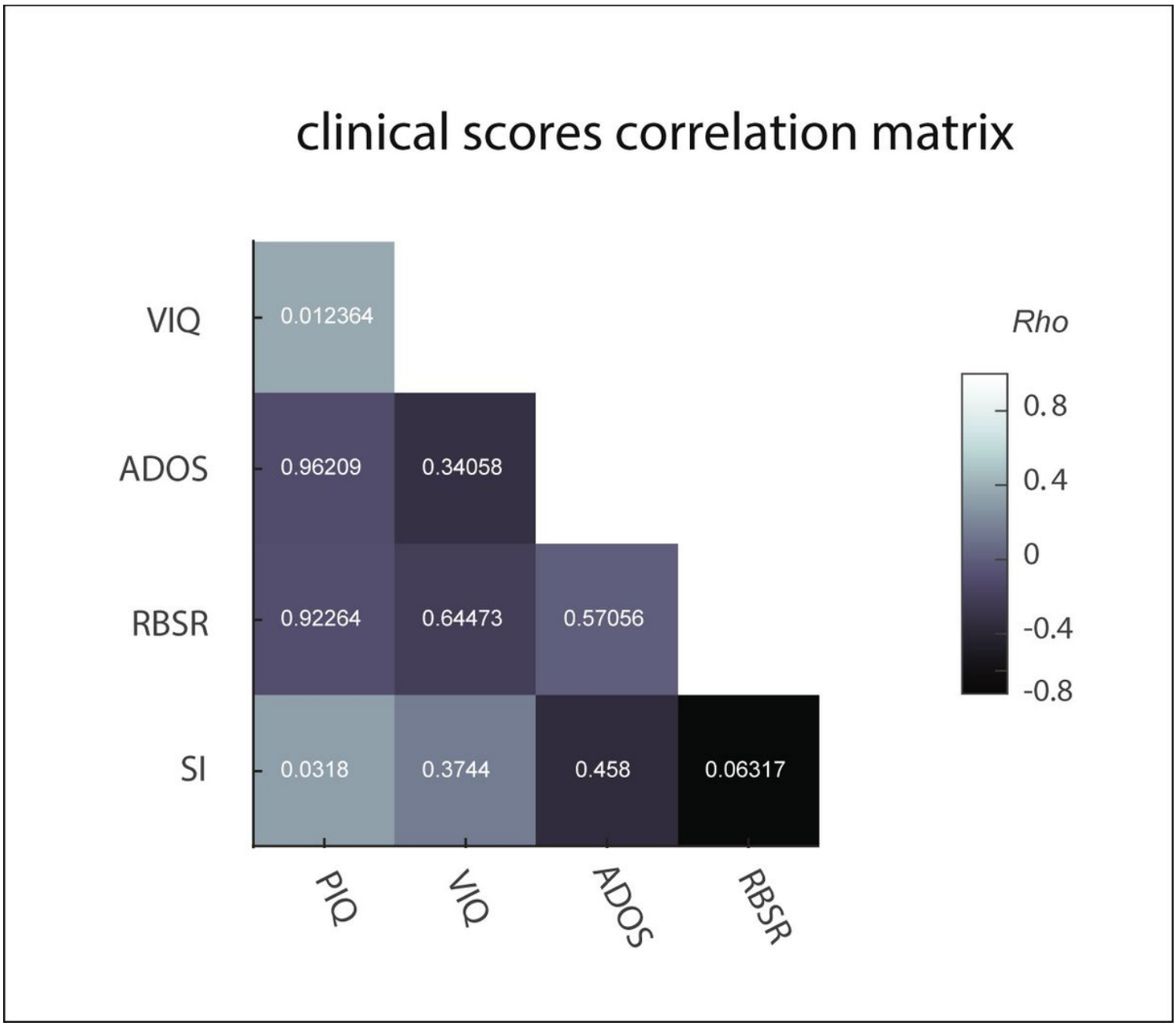

Figure 6

Correlation matrix of clinical scores and Similarity Index (SI) Gray scale colors code for Pearson rho. Uncorrected P values are given for each correlation.