



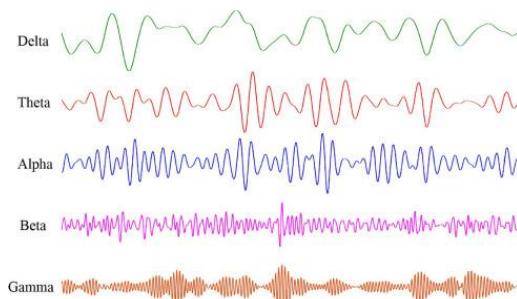
I. Self & Lab Introduction

II. Previous work

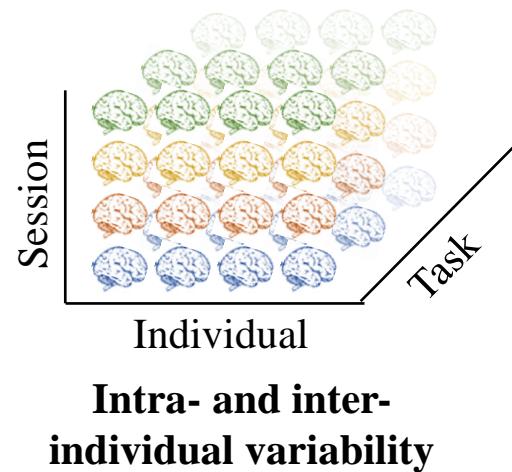
III. Lullabyte

Self-Introduction

- **Name:** Zhenxing Hu
- **Interests:** violin, badminton & nature walks.
- **Education:**
 - B.S., IOT Engineering, Xi'an University of Technology (2015-2019).
 - M.S., Biomedical Engineering, Shenzhen University (2019-2022).
- **Previous Lab:** Laboratory of Medical Informatics & Neural Dynamics
- **Award and project:**
 - The SZU outstanding graduate award for the year 2022.
 - Auditory high entropy response based research on closed-loop brain computer interface (NSFC (China) (61701316), PI: Gan Huang).



MEG/EEG



Brain-Computer Interface





Prof. Gan Huang



cofounder

Prof. Zhen Liang



Prof. Li Zhang



Dr. linling Li



Prof. Zhiguo Zhang



Founder & Former PI



MINDLAB
@szu



The MIND LAB is dedicated to developing new personalised brain-computer interaction, precise neuromodulation technologies, and intelligent learning methods for processing neuroimages and signals (including brain electricity, magnetic resonance images, genes, etc.). We are also actively applying our research results to decoding mental states, identifying neural markers, and developing new technologies for the diagnosis and treatment of brain diseases.



1. Rushuang Zhou#, Zhiguo Zhang#, Hong Fu, Li Zhang, Linling Li, Gan Huang, Fali Li, Xin Yang, Yining Dong, Yuan-Ting Zhang, and Zhen Liang*, PR-PL: A Novel Prototypical Representation based Pairwise Learning Framework for Emotion Recognition Using EEG Signals, **IEEE Transactions on Affective Computing**, 2023. (# co-first author, *co-corresponding author).
2. S. Xu#, Z. Zhang#, L. Li#, Y. Zhou, D. Lin, M. Zhang, L. Zhang, G. Huang, X. Liu, B. Becker*, and Z. Liang*, Functional Connectivity Profiles of the Default Mode and Visual Networks Reflect Temporal Accumulative Effects of Sustained Naturalistic Emotional Experience, **NeuroImage**, p.119941, 2023. (# co-first author, *co-corresponding author).
3. L. Zhang, Y. Pan, G. Huang, Z. Liang, L. Li, M. Zhang, Z. Zhang, A brain-wide genome-wide association study of candidate quantitative trait loci associated with structural and functional phenotypes of pain sensitivity, **Cerebral Cortex**, 2023.
4. Z. Liang, X. Zhang, R. Zhou, L. Zhang, L. Li, G. Huang, and Z. Zhang, Cross-Individual Affective Detection Using EEG Signals with Audio-Visual Embedding, **Neurocomputing**, 510, pp.107-121, 2022.
5. W. Hu#, Z. Zhang#, H. Zhao, L. Zhang, L. Li, G. Huang, and Z. Liang*, EEG Microstate Correlates of Emotion Dynamics and Stimulation Content during Video Watching, **Cerebral Cortex**, bhac082, pp. 1-20, 2022. (# co-first author, *corresponding author).
6. L. Zhang, G. Huang, Z. Liang, L. Li, and Z. Zhang, Estimating scale-free dynamic effective connectivity networks from fMRI using group-wise spatial-temporal regularizations, **Neurocomputing**, 485, pp. 22-35, 2022.
7. Z. Hu#, Z. Zhang#, Z. Liang, L. Zhang, L. Li, G. Huang, A New Perspective on Individual Reliability beyond Group Effects for Event-related Potentials: A Multisensory Investigation and Computational Modeling, **NeuroImage**, 250(118937), pp. 1-13, 2022. (# co-first author)
8. Z. Liang, R. Zhou, L. Zhang, L. Li, G. Huang, Z. Zhang, and S. Ishii, EEGFuseNet: Hybrid Unsupervised Deep Feature Characterization and Fusion for High-Dimensional EEG With an Application to Emotion Recognition, **IEEE Transactions on Neural Systems and Rehabilitation Engineering**, 29, pp. 1913-1925, 2021.
9. L. Li, D. Xin, H. Zhang, G. Huang, L. Zhang, Z. Liang, and Z. Zhang, Characterization of Whole-Brain Task-Modulated Functional Connectivity in Response to Nociceptive Pain: A Multisensory Comparison Study, **Human Brain Mapping**, 43, pp. 1061-1075, 2022.
10. M. Du, L. Zhang, L. Li, E. Ji, X. Han, G. Huang, Z. Liang, L. Shi, H. Yang, and Z. Zhang, Abnormal transitions of dynamic functional connectivity states in bipolar disorder: A whole-brain resting-state fMRI study, **Journal of Affective Disorders**, 289, pp. 7-15, 2021.
11. L. Zhang, Z. Fu, W. Zhang, G. Huang, Z. Liang, L. Li, B.B. Biswal, V.D. Calhoun, and Z. Zhang, Accessing Dynamic Functional Connectivity Using l0-Regularized Sparse-Smooth Inverse Covariance Estimation from fMRI, **Neurocomputing**, 443, pp. 147-161, 2021.
12. Z. Liang, F. Li, W. Hu, G. Huang, S. Oba, Z. Zhang, and S. Ishii, A Generalized Encoding System for Alpha Oscillations through Visual Saliency Analysis, **IEEE Transactions on Neural Systems and Rehabilitation Engineering**, 28(12), pp. 2731-2743, 2020.
13. Q. Lin, G. Huang, L. Li, L. Zhang, Z. Liang, A.M. Anter, and Z. Zhang, Designing individual-specific and trial-specific models to accurately predict the intensity of nociceptive pain from single-trial fMRI responses, **NeuroImage**, 225(117506), pp. 1-15, 2020.
14. A.M. Anter, G. Huang, L. Li, L. Zhang, Z. Liang, and Z. Zhang, A New Type of Fuzzy-Rule-Based System With Chaotic Swarm Intelligence for Multiclassification of Pain Perception From fMRI, **IEEE Transactions on Fuzzy Systems**, 28(6), pp.1096-1109, 2020.



I. Self & Lab Introduction

II. Previous work

III. Lullabyte



I. **ERP reliability analysis**

(A multisensory investigation and computational modeling of ERP reliability)

Time: 2020-2021

II. **EEG-based biometric challenge**

(A multi-subject, multi-session, and multi-task database for EEG-based biometrics challenge)

Time: 2021-2022

III. **Theta-wave modulation by auditory stimuli**

(Auditory high entropy response based closed-loop brain computer interface)

Time: 2022-present

I. ERP reliability analysis

Reliability in clinical research

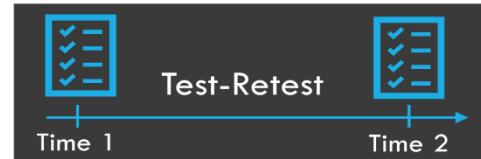
Reliability:

- The ability of a measure to give **consistent** results under repeated tests.

1. Inter-rater reliability



2. Test-retest reliability



3. Internal consistency



Test-retest reliability in psycho-diagnosis:

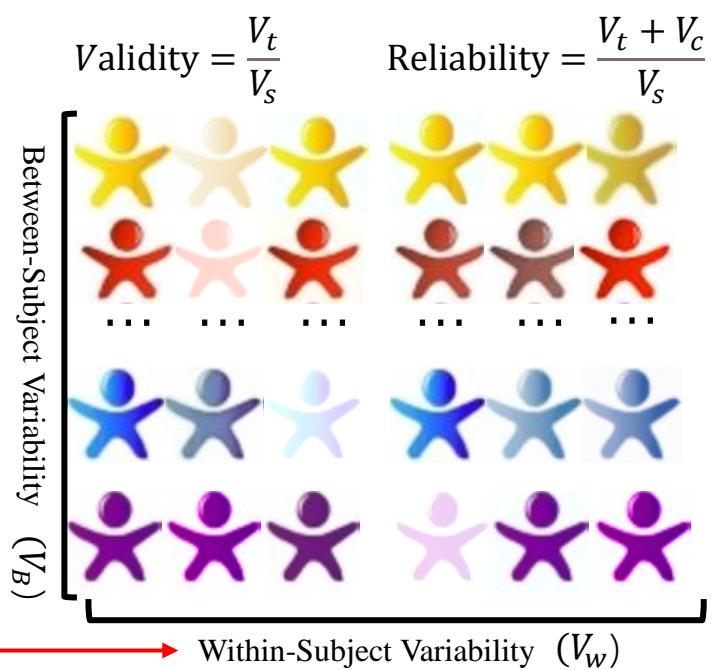
- The ability of a measure to yield a **reproducible** difference between individuals

$$\text{Total} \\ V_s = V_t + V_c + V_e$$

$$\text{Disorder} \\ V_t$$

$$\text{Contaminants} \\ V_c$$

$$\text{Random Error} \\ V_e$$

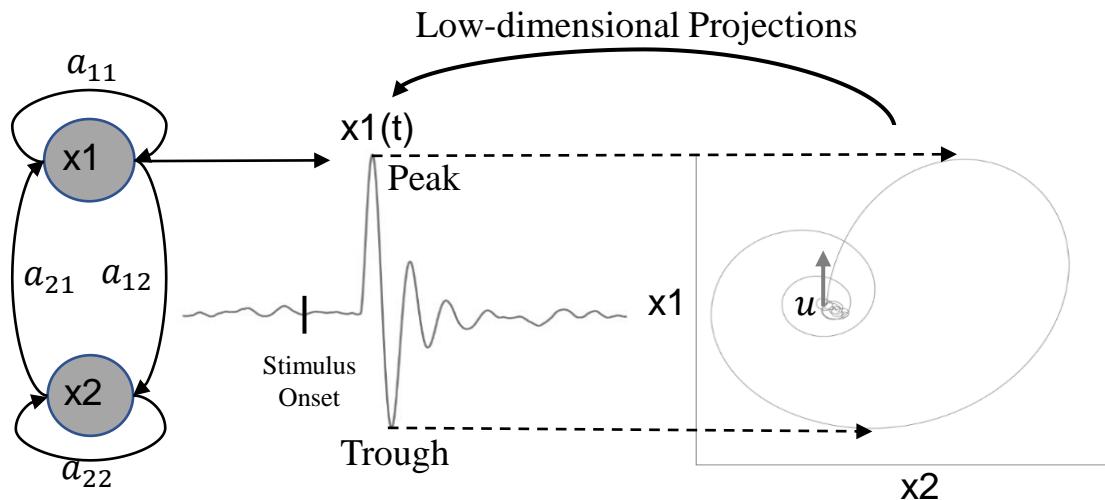


I. ERP reliability analysis

Motivations

“The current analysis about **ERP reliability** only focus on narrow window around the peaks”^[1]

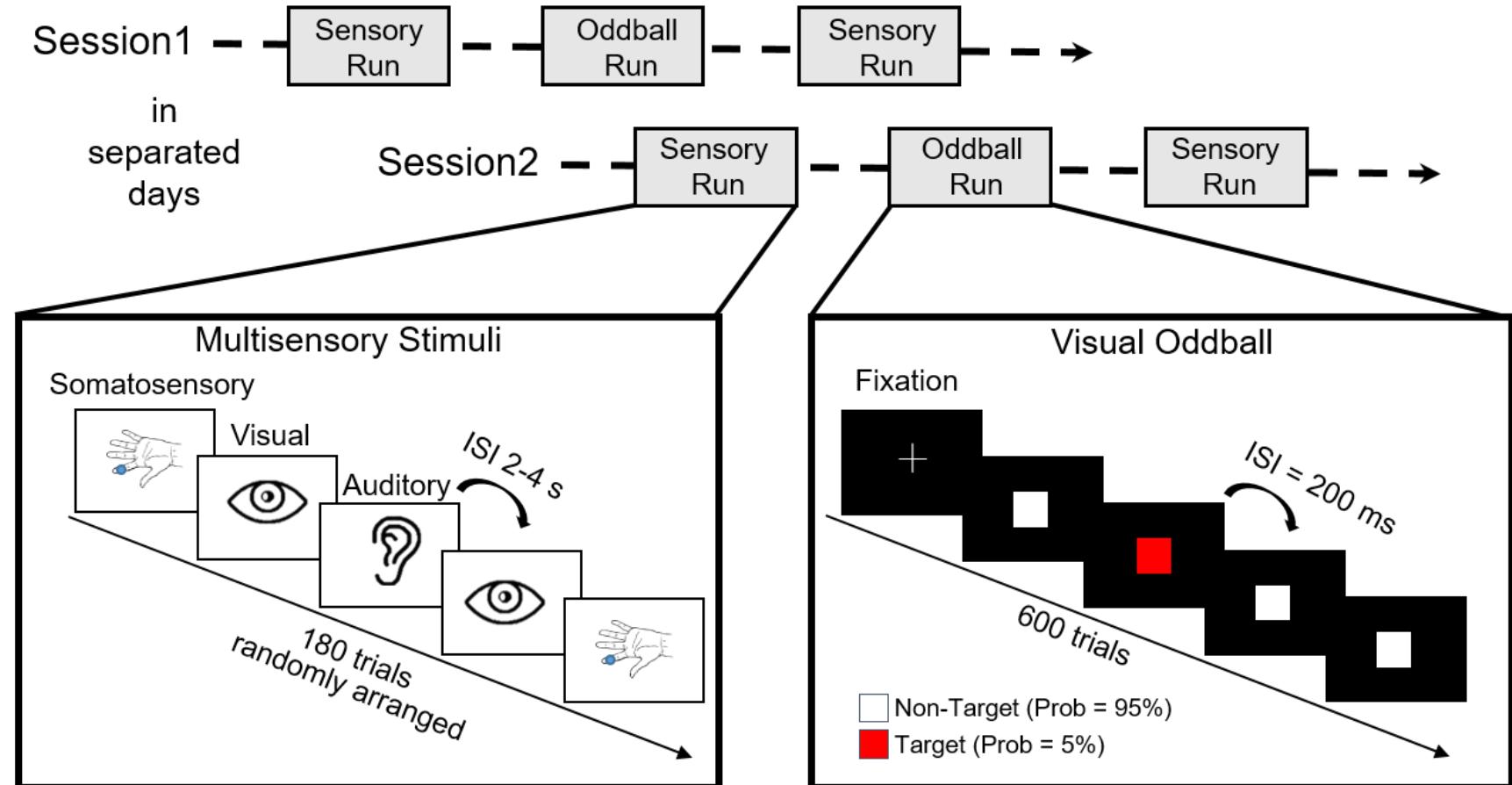
- Computational models of neural processes rarely considered when a process “peaks”.
- Both peaks and zero-crossing point of ERPs just represent different stages of one unified neurophysiological process.



[1] Gaspar, Carl M., et al. "Reliability of ERP and single-trial analyses." *Neuroimage* 58.2 (2011): 620-629.

I. ERP reliability analysis

➤ Part I. ERP reliability analysis



Hu, Z., et al., 2022. A new perspective on individual reliability beyond group effect for event-related potentials: A multisensory investigation and computational modeling. *NeuroImage*, 250, p.118937.

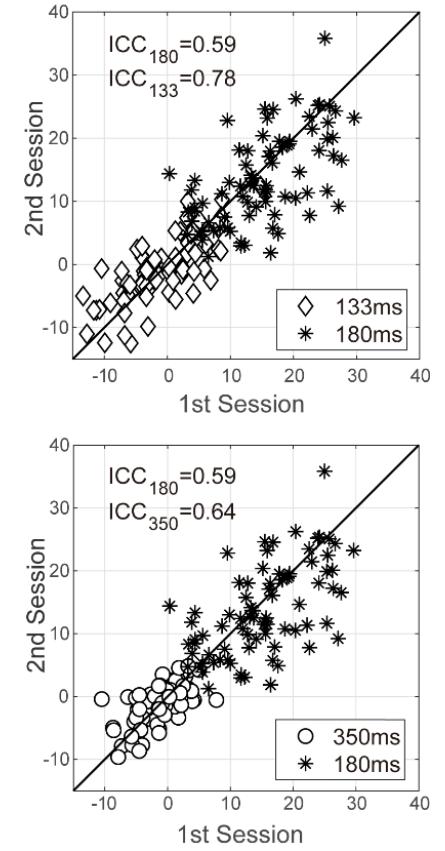
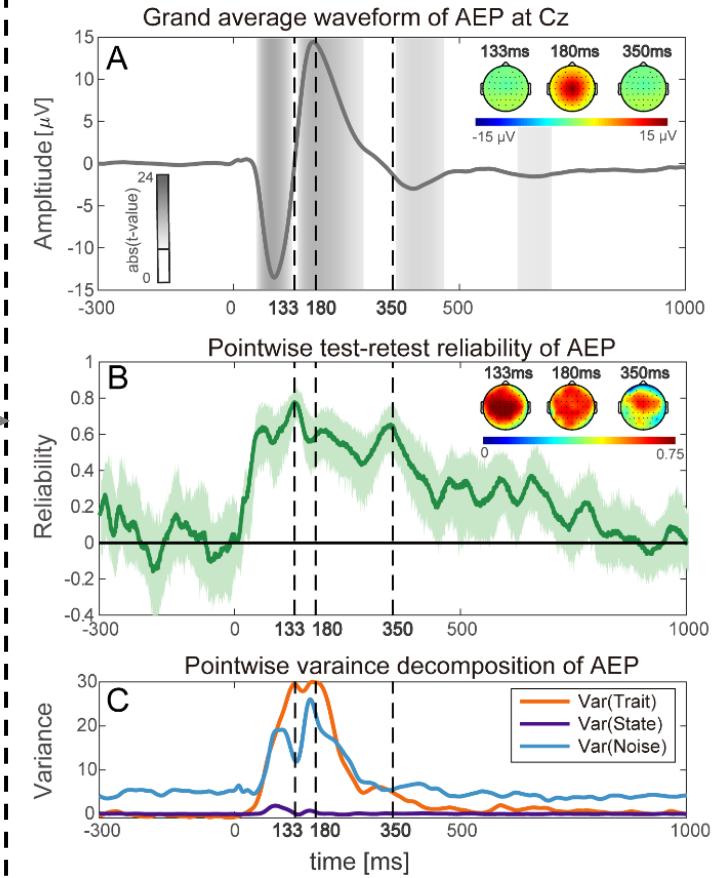
I. ERP reliability analysis

$$ERP_{ij} = \mu + r_i + C_j + rc_{ij} + e_{ij}$$

$$\begin{array}{ccc|c} r_1 & r_1 & r_1 \\ r_2 & r_2 & r_2 \\ r_3 & r_3 & r_3 \\ \hline \bar{r} & \bar{r} & \bar{r} \end{array} \rightarrow Var(Trait)$$

$$\begin{array}{ccc|c} c_1 & c_2 & \bar{c} \\ c_1 & c_2 & \bar{c} \\ c_1 & c_2 & \bar{c} \\ \hline c_1 & c_2 & \bar{c} \end{array} \rightarrow Var(State)$$

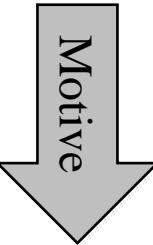
$$\begin{array}{ccc|c} e_{11} & e_{12} & \bar{e}_{1r} \\ e_{21} & e_{22} & \bar{e}_{2r} \\ e_{31} & e_{32} & \bar{e}_{3r} \\ \hline \bar{e}_{c1} & \bar{e}_{c2} & \bar{e} \end{array} \rightarrow Var(Noise)$$



Conclusion: Stronger response amplitude doesn't guarantee higher reliability

I. ERP reliability analysis

Statistical Analysis



Motive

ERPs	abs(<i>t</i> -value)		Hilbert envelope		between-subject variance	
	Spearman's ρ	<i>p</i> -value	Spearman's ρ	<i>p</i> -value	Spearman's ρ	<i>p</i> -value
AEP (Cz)	-0.19	3.04×10^{-5}	0.27	$< 10^{-12}$	0.36	$< 10^{-12}$
SEP (Cz)	0.38	$< 10^{-12}$	0.51	$< 10^{-12}$	0.71	$< 10^{-12}$
VEP (Oz)	0.17	0.002	0.12	0.017	0.75	$< 10^{-12}$
P300 (Pz)	0.54	$< 10^{-12}$	0.74	$< 10^{-12}$	0.84	$< 10^{-12}$

Mathematical Model

$$x'(t) = Ax(t) + C * u(t) + e(t)$$

$$A = \begin{bmatrix} -c & d \\ -d & -c \end{bmatrix}$$

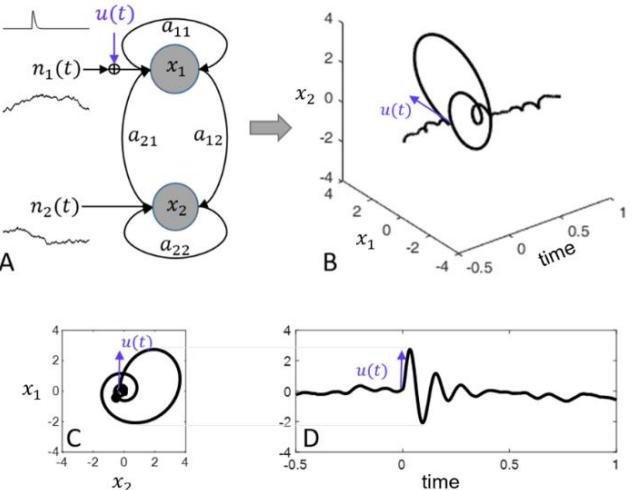
$$x = [x_1, x_2]$$

$$C = C_{sub} + C_{trial} \in (\mu_{trial}, \sigma_{trial}^2)$$

$$u(t) = \begin{cases} at * e^{-bt} & t \geq jitter_{sub} \\ 0 & t < jitter_{sub} \end{cases}$$

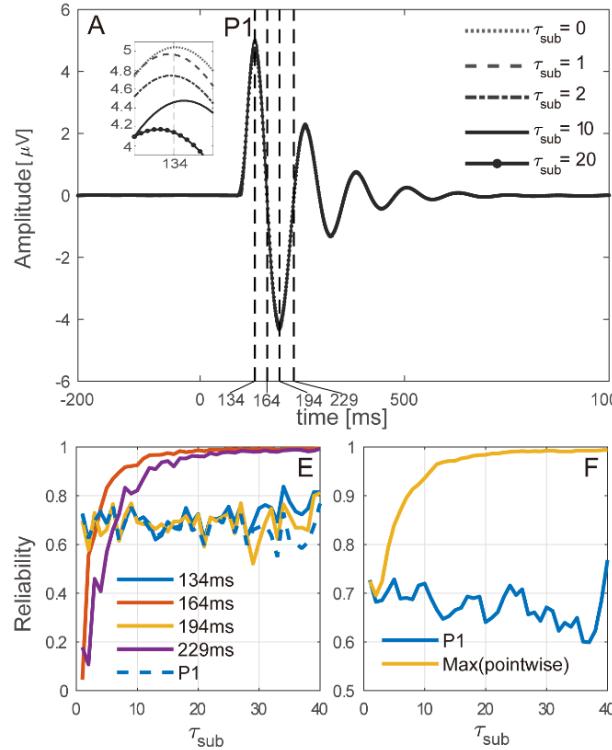
$$jitter_{sub} \in (-\tau_{sub}, \tau_{sub})$$

Model Diagram

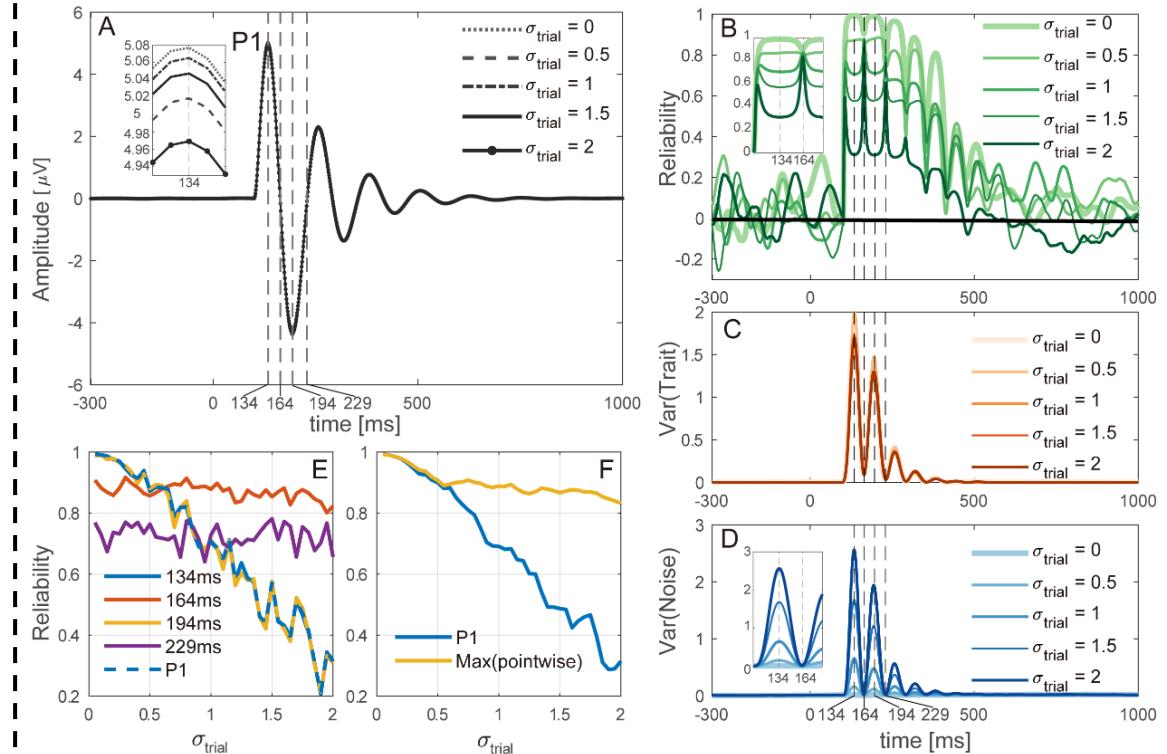


I. ERP reliability analysis

The influence of trial-level variability: τ_{sub}



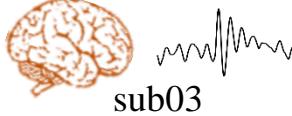
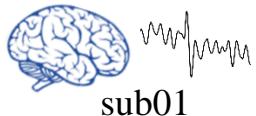
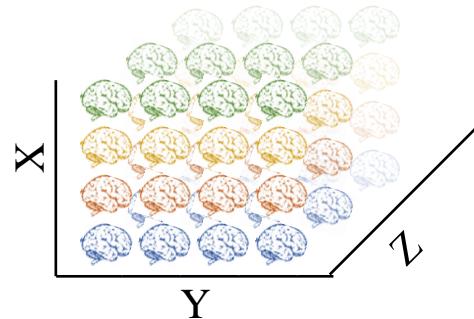
The influence of trial-level variability: σ^2_{trial}



Conclusion: τ_{sub} and σ^2_{trial} can influence the consistency between group effects and individual reliability in different ways

II. EEG-based biometrics challenge

EEG signals are modulated by multiple variables
(Subjects, Task, Session, Methodology,...)



Inter-subject variability of EEG spatiotemporal patterns



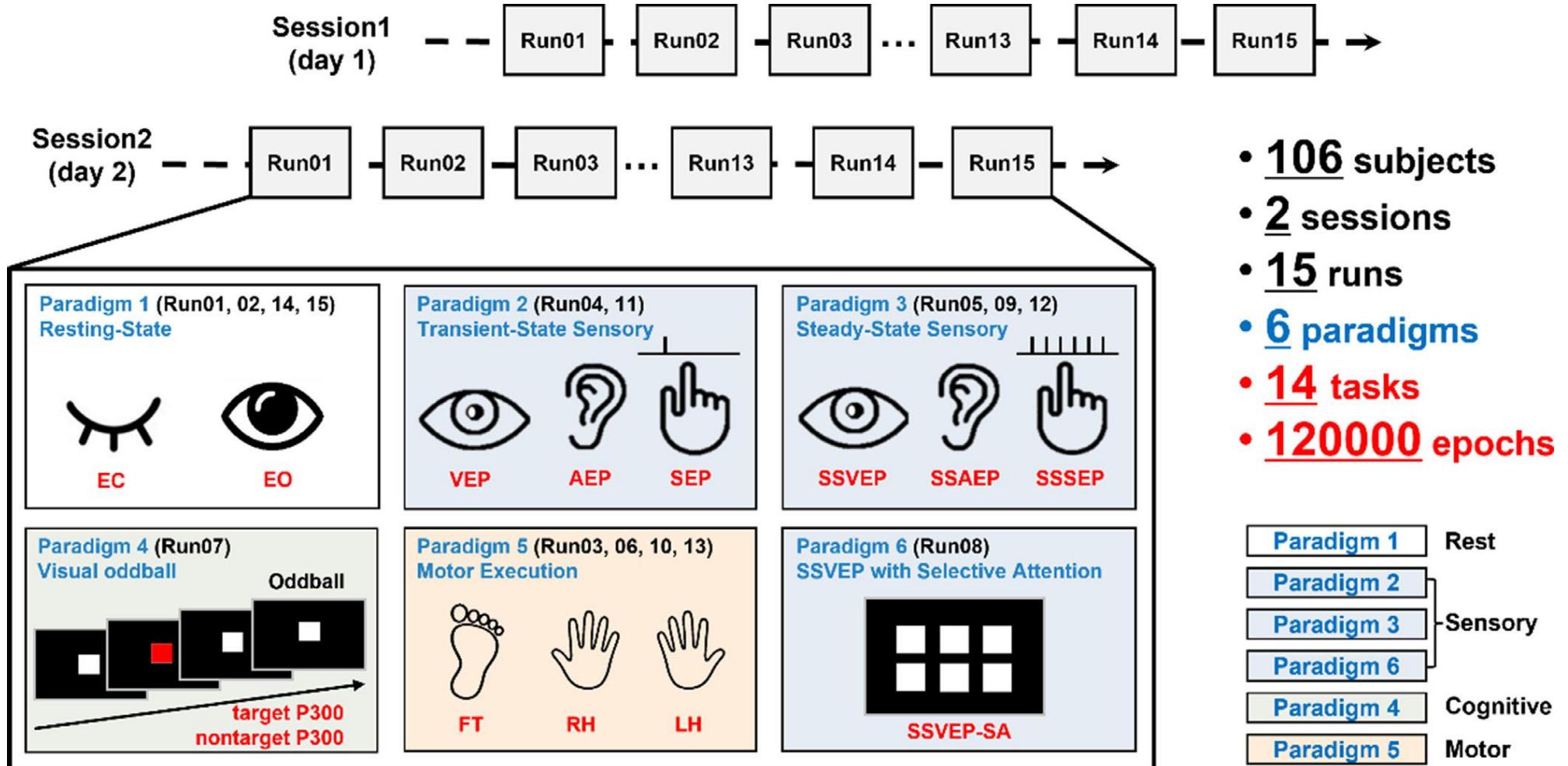
Intra-subject variability over different tasks and sessions



Research Question

Can **subject identity** be robustly decoded from EEG signals
across different **tasks** and **sessions**?

II. EEG-based biometrics challenge



II. EEG-based biometrics challenge

➤ Motivations

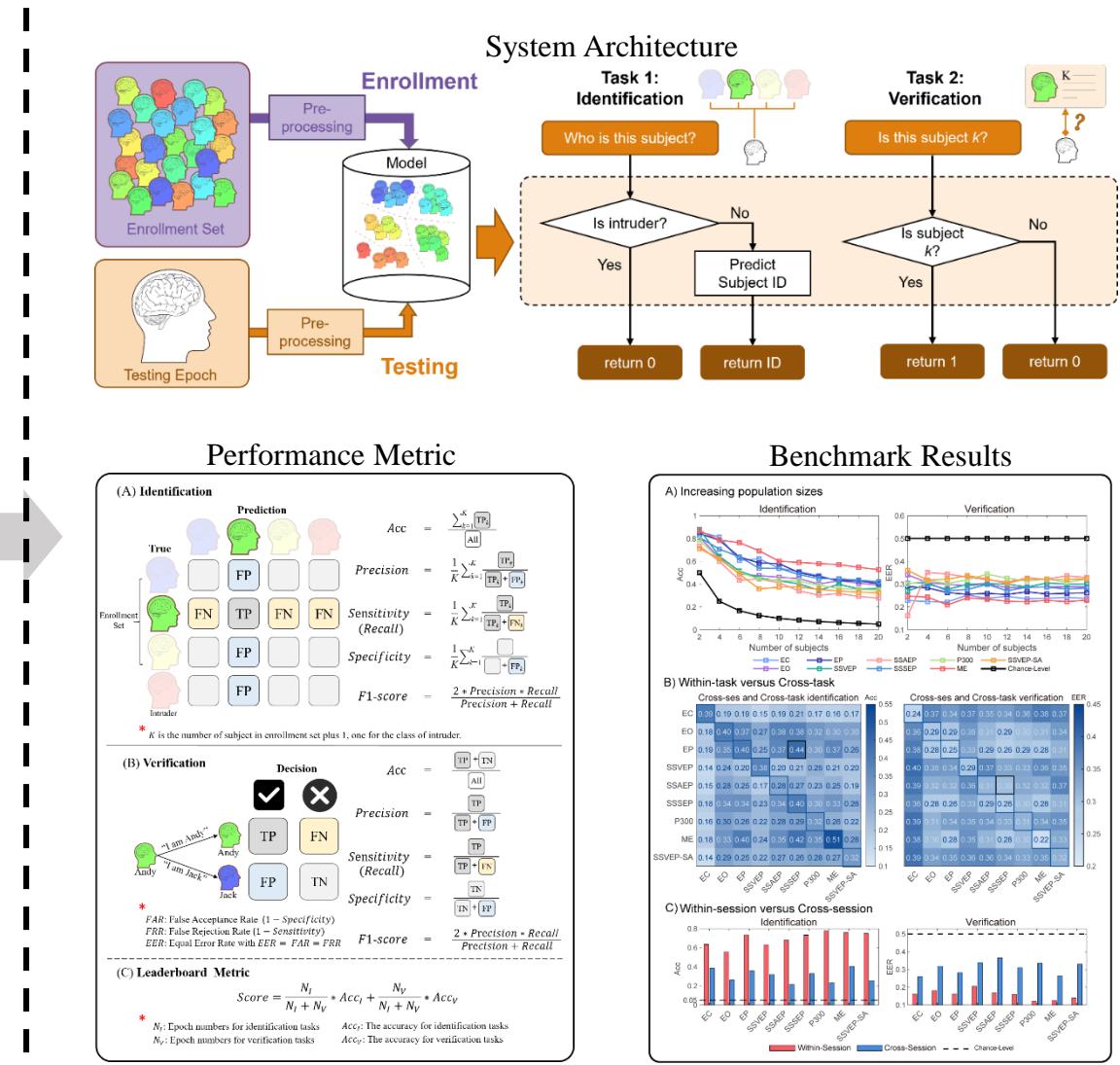
- The lack of a unified test benchmark and platform.
- The lack of A Multi-subject, Multi-session, and Multi-task database.

➤ Goal

- Advancing the development of machine learning algorithms for EEG decoding.
- Deepening the understanding of intra- and inter-subject variability of EEG signals.

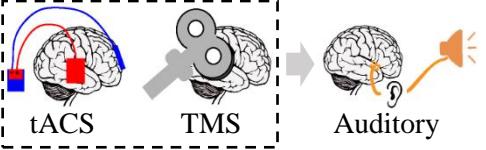
➤ My work

- Figure visualization.
- Benchmark code writing.
- Writing – original draft.

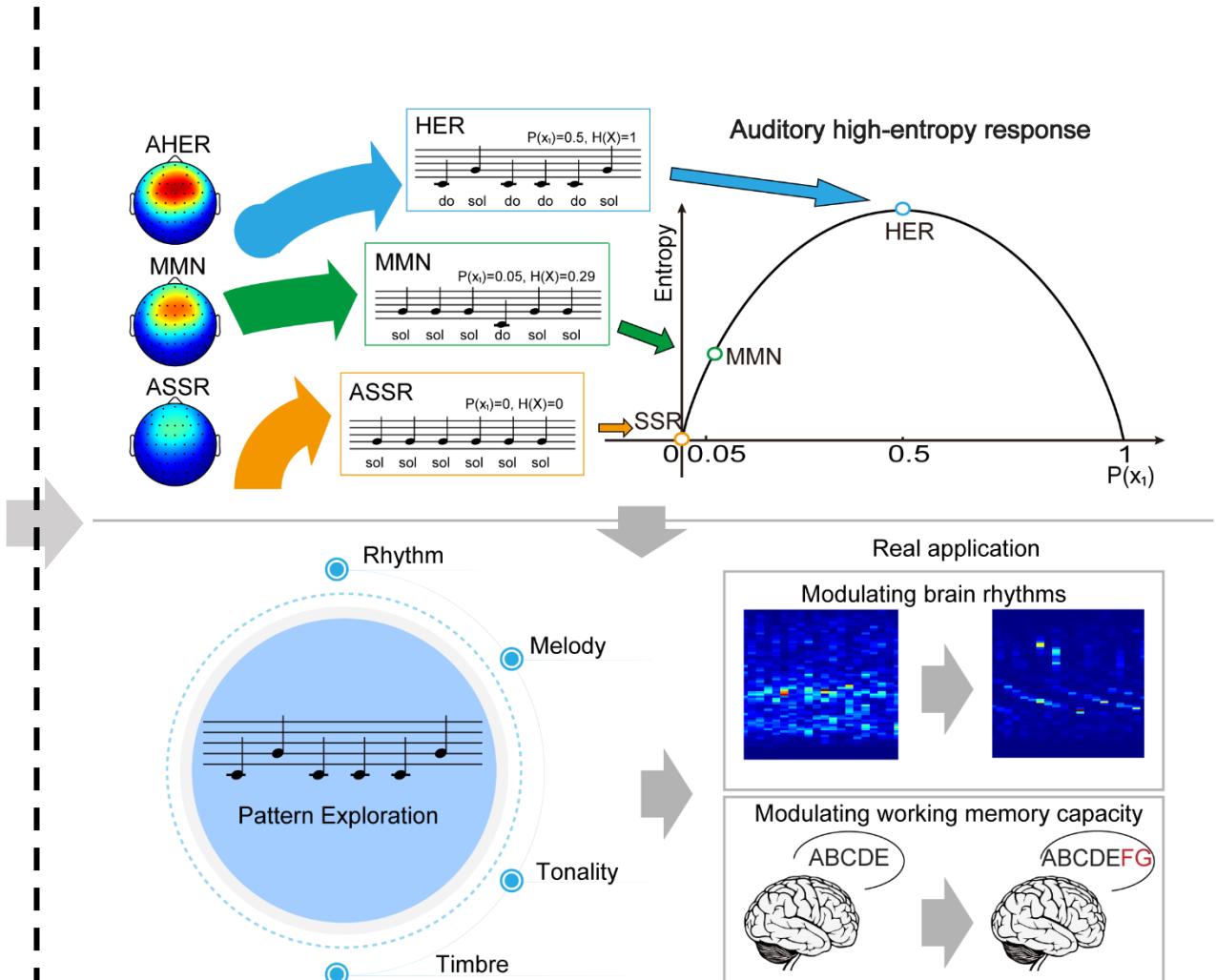


III. Brain-wave modulation by auditory stimuli

Background

- NIBS (Non-Invasive Brain Stimulation)
 
- Current limitation
 - High certainty leads to repetition suppression.
 - Non-repetitive small probability events in MMN cannot modulate brain rhythm.

- Research Question
 - The influences of acoustic sequence entropy on prefrontal theta-wave modulation
- My Work
 - Experimental design
 - EEG collection and analysis

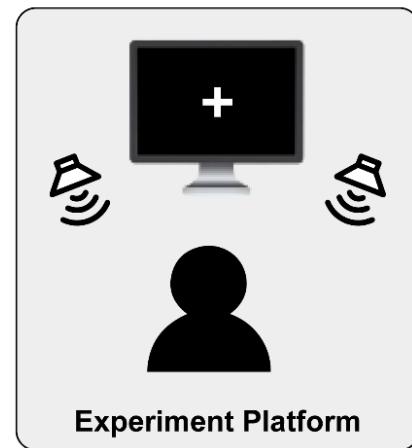
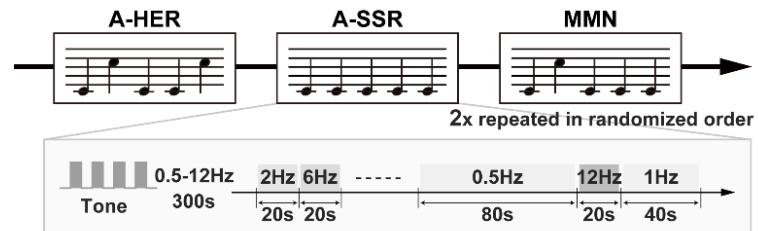


Toussaint, G.T., 2019. The geometry of musical rhythm: what makes a "good" rhythm good?. CRC Press.
 Gündüz, G., 2023. Entropy, energy, and instability in music. *Physica A: Statistical Mechanics and its Applications*, 609, p.128365.

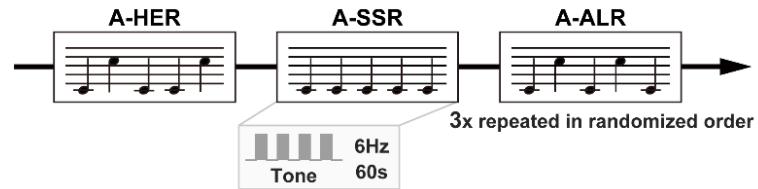
III. Theta-wave modulation by auditory stimuli

Experimental Paradigm

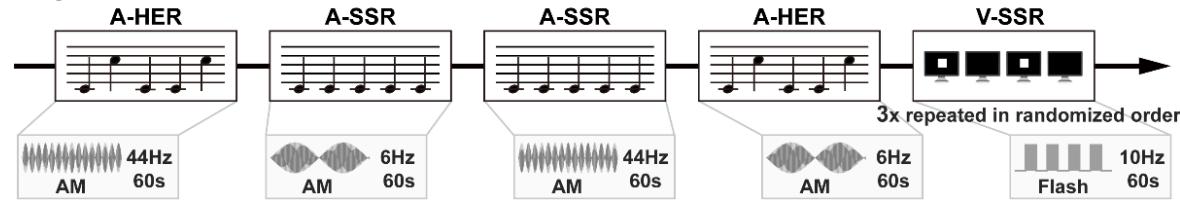
Experiment 1:



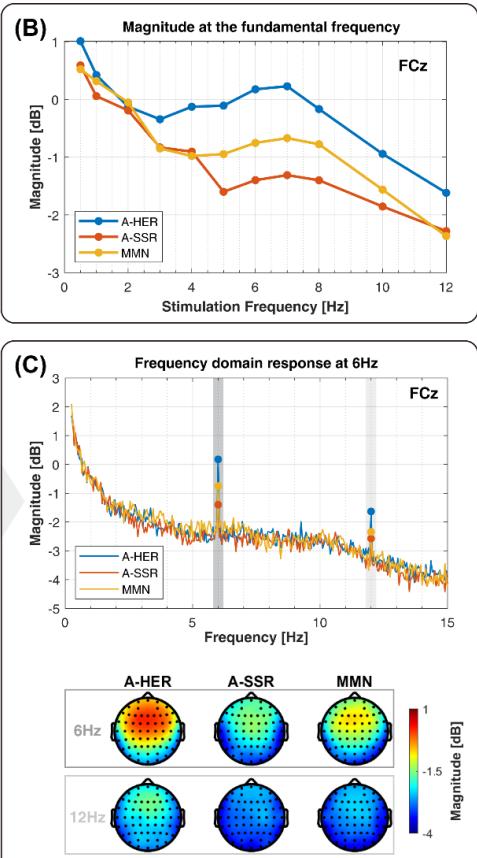
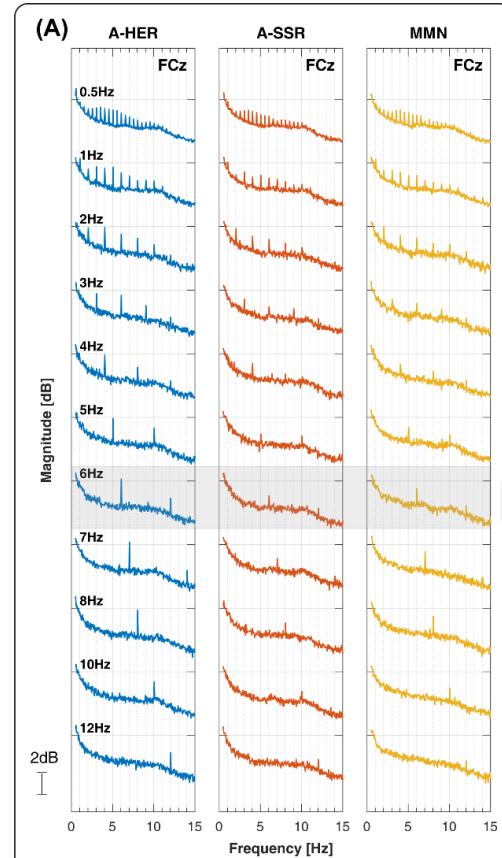
Experiment 2:



Experiment 3:

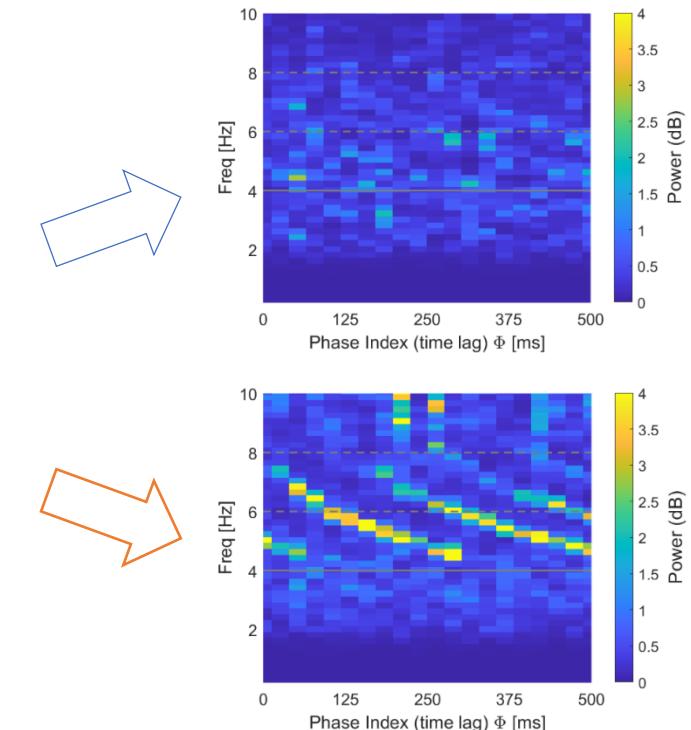
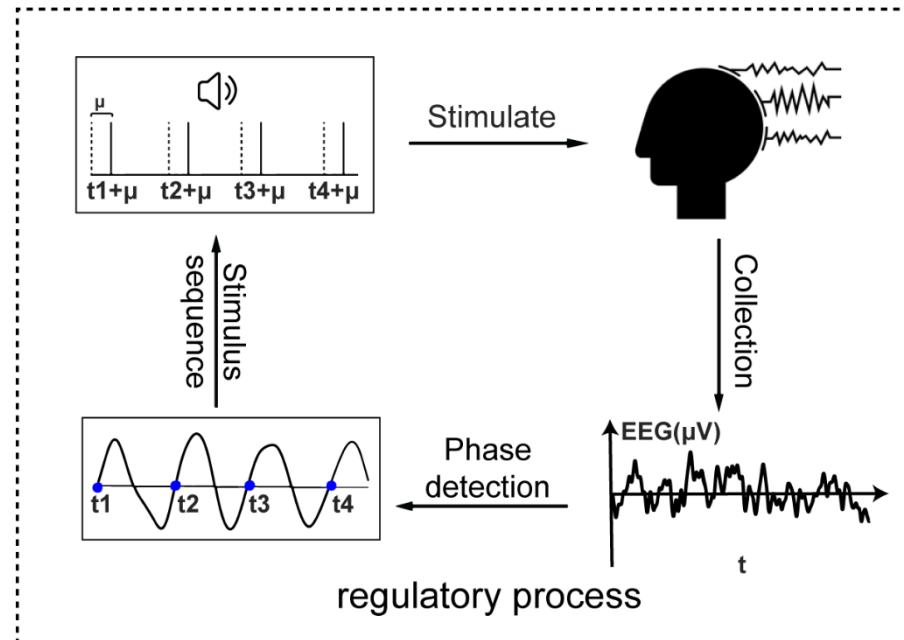
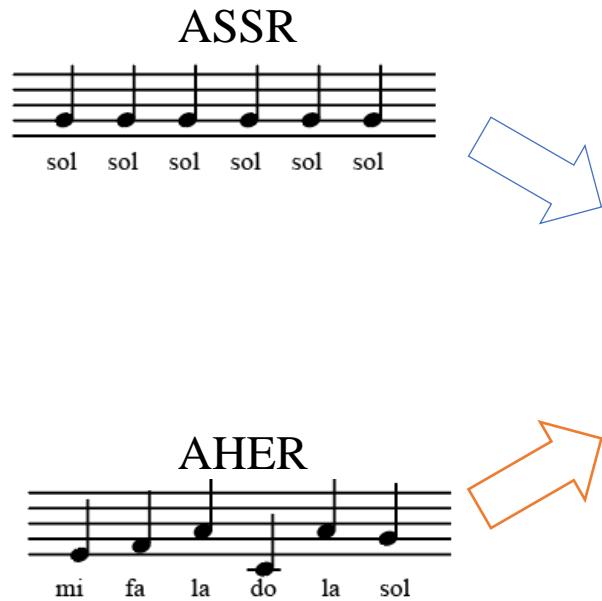


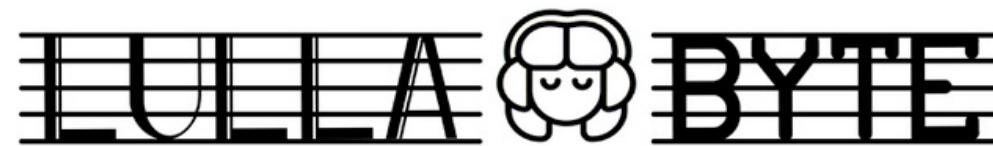
Results



III. Theta-wave modulation by auditory stimuli

Phase-locked auditory Stimulation





Unraveling the Effects of Music on Sleep through Musicology, Neuroscience, Psychology and Computer Science



UNIVERSITÉ DE FRIBOURG
UNIVERSITÄT FREIBURG



Radboudumc

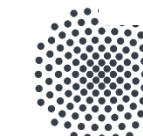
TECHNISCHE
UNIVERSITÄT
DRESDEN



AARHUS
UNIVERSITY



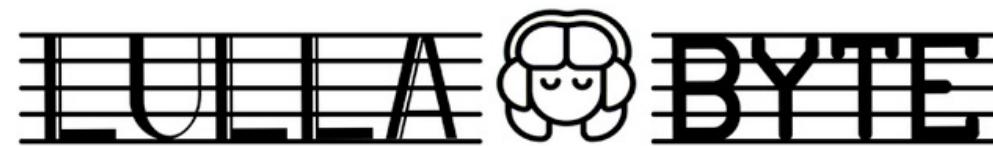
Universitat
Pompeu Fabra
Barcelona
INSTITUT DE RECHERCHE



University of Stuttgart
Germany



femto-st
SCIENCES &
TECHNOLOGIES



Unraveling the Effects of Music on Sleep through Musicology, Neuroscience, Psychology and Computer Science



UNIVERSITÉ DE FRIBOURG
UNIVERSITÄT FREIBURG



Radboudumc

TECHNISCHE
UNIVERSITÄT
DRESDEN



AARHUS
UNIVERSITY



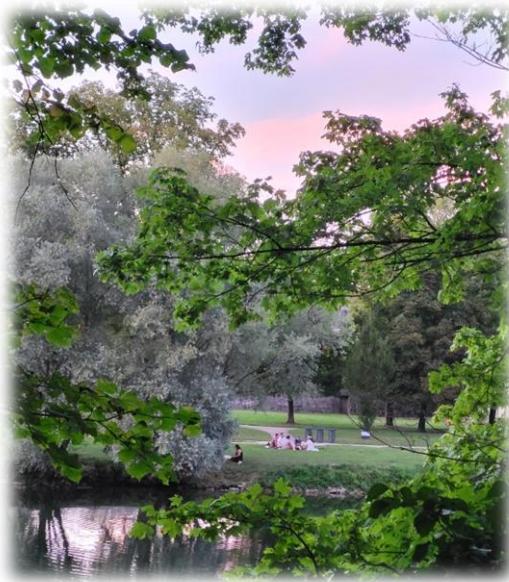
Universitat
Pompeu Fabra
Barcelona
INSTITUT DE RECHERCHE



University of Stuttgart
Germany

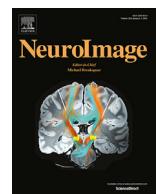


femto-st
SCIENCES &
TECHNOLOGIES





Thanks!



A new perspective on individual reliability beyond group effect for event-related potentials: A multisensory investigation and computational modeling



Zhenxing Hu^{a,b,1}, Zhiguo Zhang^{a,b,c,d,1}, Zhen Liang^{a,b}, Li Zhang^{a,b}, Linling Li^{a,b}, Gan Huang^{a,b,*}

^a School of Biomedical Engineering, Health Science Center, Shenzhen University, Shenzhen, Guangdong, 518060, China

^b Guangdong Provincial Key Laboratory of Biomedical Measurements and Ultrasound Imaging, Shenzhen University, Shenzhen, Guangdong, 518060, China

^d Marshall Laboratory of Biomedical Engineering, Shenzhen University, Shenzhen, 518060, China

^c Peng Cheng Laboratory, Shenzhen, Guangdong, 518055, China

ARTICLE INFO

Keywords:

Event-related potentials
Multisensory
Group effects
Individual reliability
Computational Model

ABSTRACT

The dominant approach in investigating the individual reliability for event-related potentials (ERPs) is to extract peak-related features at electrodes showing the strongest group effects. Such a peak-based approach implicitly assumes ERP components showing a stronger group effect are also more reliable, but this assumption has not been substantially validated and few studies have investigated the reliability of ERPs beyond peaks. In this study, we performed a rigorous evaluation of the test-retest reliability of ERPs collected in a multisensory and cognitive experiment from 82 healthy adolescents, each having two sessions. By comparing group effects and individual reliability, we found that a stronger group-level response in ERPs did not guarantee higher reliability. A perspective of neural oscillation should be adopted for the analysis of reliability. Further, by simulating ERPs with an oscillation-based computational model, we found that the consistency between group-level ERP responses and individual reliability was modulated by inter-subject latency jitter and inter-trial variability. The current findings suggest that the conventional peak-based approach may underestimate the individual reliability in ERPs and a neural oscillation perspective on ERP reliability should be considered. Hence, a comprehensive evaluation of the reliability of ERP measurements should be considered in individual-level neurophysiological trait evaluation and psychiatric disorder diagnosis.

1. Introduction

Event-related potentials (ERPs) are noninvasive electrophysiological measures of indexing a range of sensory, cognitive, and motor processes involved in human brain activity. In clinical and translational applications of ERPs, a key challenge is to identify a reliable and valid mapping between individuals' brain activation and their perceptual or cognitive capacities (Nelson and Guyer, 2012). Measurement reliability is the prerequisite for clinical applications of ERPs, such as assessments of meditative practice using sensory-evoked potentials (Cahn and Polich, 2006) or diagnoses of psychiatric cognitive dysfunction by cognitive ERPs like P300 (Polich, 2004), and studies concerning reliability have received more attention recently (Dubois and Adolphs, 2016; Höller et al., 2017; Noble et al., 2019; Croce et al., 2020).

Originating from the field of psychometrics, reliability reflects the "trustworthiness" of a measure and denotes the extent to which a measure will yield a reproducible difference between individuals

(Kraemer, 2014). The importance of reliability in the research of individual difference cannot be overstated, regardless of the data analytics approaches used (e.g., correlational analysis or machine learning). In correlational analysis, the ability to find correlations between brain activation and cognitive behavior depends on the reliability of these measures (Goodhew and Edwards, 2019). In other words, the maximum possible correlation is constrained by the reliability of the individual measures used to calculate the correlation (Spearman, 1910). In machine learning-based individualized prediction, reliability has been proved mathematically to provide a lower bound on predictive accuracy (Bridgeford et al., 2020).

Since the first systematic study on the reliability of ERPs (Segalowitz and Barnes, 1993), numerous studies have evaluated the test-retest reliability of ERP amplitude and the latency elicited from a variety of experimental paradigms (Cassidy et al., 2012; Cruse et al., 2014), but the primary focus has always been restricted to narrow time windows around ERP peaks (Thigpen et al., 2017; Cruse et al., 2014; Ip et al.,

* Corresponding author at: School of Biomedical Engineering, Health Science Center, Shenzhen University, Shenzhen, 518060, China.

E-mail address: huanggan1982@gmail.com (G. Huang).

¹ Zhenxing Hu and Zhiguo Zhang contributed equally to this paper.

2018). Characteristic features, including latency, maximum amplitude, mean amplitude, and area under the window, are typically used to examine the reliability of ERPs. These ERP features are used in a machine learning model or correlation analysis to establish linkage between ERPs and cognitive/behavioral variables (Hu and Iannetti, 2019). However, such an analysis routine implicitly assumes that only the peak-related ERP measures reflect the subject-specific neurophysiological process to an external stimulus. This assumption is problematic because the entire ERP shapes (rather than latency and amplitude of ERP peaks) are physiologically meaningful and important (Gaspar et al., 2011). Taking the temporal evolution of facial emotion perception as an example, the temporal shape of ERP can provide valuable clues about processing dynamics beyond what can be inferred from data restricted to ERP peaks (Van Rijssbergen and Schyns, 2009).

ERP peaks represent the strongest group effects (i.e., group-level experimental effects among different conditions/cohorts). More specifically, by comparing the ERP response with its baseline activity, or contrasting two experimental conditions (i.e., the ERP difference wave), peak-related features of well-known ERP components, like N100, N200, and P300, were claimed to be closely associated with various perceptual and cognitive variables (Sur and Sinha, 2009). Here, the focus was on significant group effects responding to one condition versus another. As a representative example relevant to this research, the P300 was found to reflect the processes involved in stimulus evaluation or categorization as evidenced by experimental manipulation; thus, it is often reasonable to ask whether peak-related features of P300 reflect an individual's cognitive function. From this perspective, as the dominant approach in investigating individual differences in ERPs, peak-based analysis implicitly employs group-level prior information. However, from the perspective of individual difference, it remains unclear whether peak-related activity shows robustness or consistency in assessing between-subject variance (Brandmaier et al., 2018).

Indeed, the approach of identifying regions-of-interests (ROIs) by the strongest group effects and subsequently testing them for individual reliability was a common practice in evaluating individual differences in ERP studies, but recent studies have raised concerns that such a conventional approach may reduce the probability of detecting significant individual-level effects, especially in functional magnetic resonance imaging (fMRI) (Fröhner et al., 2019; Infantolino et al., 2018). For researchers interested in individual differences, between-subject variance in brain function is usually considered as the signal of interest rather than noise (Seghier and Price, 2018). For researchers interested in experimental effects, within-subject variance is treated as the signal of interest, and between-subject variance represents the noise that should be minimized. Those different views imply that regions eliciting greater activation (i.e., a peak at an electrode showing the strongest group-averaged activity) on group effects may not correspond to reliable individual effects, which has been thoroughly discussed in psychology recently (Hedge et al., 2018; Goodhew and Edwards, 2019; Fisher et al., 2018). To the best of our knowledge, the rationality of selecting individual difference variables based on group effects in ERP analysis has been seldom challenged. Whether and in which situation the group effects and individual reliability are consistent is still questionable. In real data, the underlying factors among different subjects are unmeasurable and cannot be adjusted at will, which makes it challenging to answer this question. Thus, a simulation model should be applied to investigate underlying factors of modulating the consistency between the group effect and individual reliability, but this investigation via computational modeling is still absent.

To address the abovementioned problems, the present study sought to examine the test-retest reliability of sensory-evoked potentials and cognitive ERPs based on the whole waveforms but not those restricted to narrow time windows around the peaks. More specifically, to test whether there is a spatial and temporal dissociation between group effects and the individual reliability result, the reliability of auditory-evoked potential (AEP), somatosensory-evoked potential (SEP), visual-

evoked potential (VEP), and P300 were systematically examined by spatiotemporal decomposition and evaluation in a pointwise way (i.e., at each spatial-temporal EEG sample). Further, a dynamical system model was applied for the simulation of ERP generation to investigate the underlying mechanism explaining the real data results, in which key model parameters were varied to test their influences on the consistency between group effects and individual reliability. Data and code are available online (<https://osf.io/v59qu>).

2. Materials and methods

2.1. Data collection and preprocessing

2.1.1. Subject information

A total of 106 healthy subjects participated in this study, and 95 subjects ($\text{Mean}_{\text{age}} = 21.3$ years; $\text{SD}_{\text{age}} = 2.2$ years, 73 males) among them attended two sessions, which were scheduled on different days, separated more than 6 days and 20 days apart on average. After removing 13 subjects whose data were corrupted with heavy artifacts, 82 subjects were included in subsequent reliability analyses. Ethical approval of the study was obtained from the Medical Ethics Committee, Health Science Center, Shenzhen University (No. 2,019,053). All subjects were informed of the experimental procedure, and they signed informed consent before the experiment.

2.1.2. Experimental paradigm

As illustrated in Fig. 1, the experimental paradigm was the same for the two sessions on different days. The experimental paradigm contained three types of sensory-evoked experiments (visual, auditory, and somatosensory) and a cognitive visual oddball experiment. Multiple sensory stimuli were arranged in two runs for each session. Each run consisted of 90 trials, including visual, auditory, and somatosensory vibration stimuli. These stimuli were delivered in a random order with inter-stimulus-interval (ISI) randomly distributed in the range of 2–4 s. Each stimulus lasted 50 ms. Hence, for each subject, there were a total of 180 trials of sensory stimulation in each session and 60 trials for each of the visual, auditory, and somatosensory stimuli. The P300 experiment was arranged between the two runs of multiple sensory stimuli for each session. The visual oddball experiment was performed with the red squares as the target stimuli and the white squares as the nontarget stimuli on the screen. Each square lasted 80 ms, with an ISI of 200 ms. Hence, a total of 600 trials were delivered within 2 min in a run, in which the target stimuli appeared with the possibility of 5%. A subject was asked to count the number of red squares and report the result at the end of the run to keep his/her attention on the screen.

2.1.3. Platform setup

During the experiment, the subjects were seated in a comfortable chair. For multisensory stimuli, an Arduino Uno platform was programmed to release the three types of stimuli, which communicated with the Matlab program (The MathWorks Inc., Natick, USA) on a PC through a serial port. Visual stimuli were delivered by a 3 W light-emitting diode (LED) with a 2 cm diameter circular light shield, which is placed 45 cm away from subjects' eyes. The LED intensity was 1074 Lux as measured by a light meter (TES-1332A, TES). Auditory stimuli were presented via a Nokia WH-102 headphone. The intensity is set at a comfortable level (75 dB SPL) for all subjects as measured by a digital sound level meter (Victor 824, Double King Industrial Holdings Co., Ltd. Shenzhen, China). Somatosensory stimuli were generated by a 1027 disk vibration motor with the rated power 3 W, efficiency 80%, and dimensions 10mm*2.7 mm). For the visual oddball P300 experiment, a 24.5-inch screen with a 240-Hz refreshing rate (Alienware AW2518H, Miami, USA) was used to present the visual stimuli. The 300×300 pixels red and white squares were delivered in sequence in the center of the 1920×1080 pixels screen with the background in black.

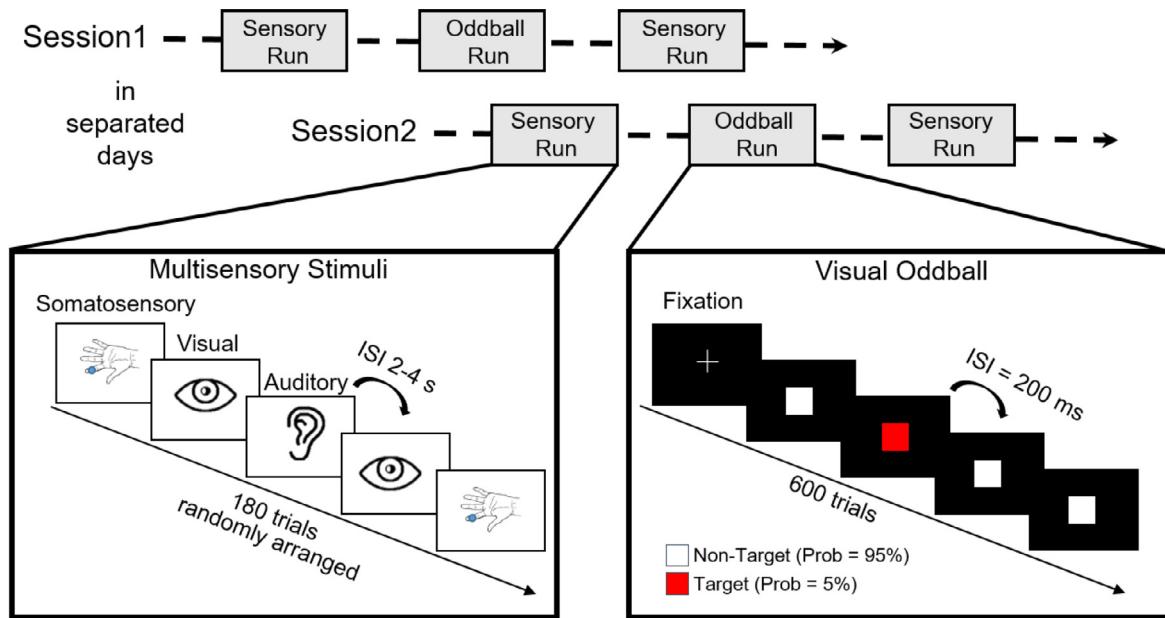


Fig. 1. Experiment procedure. Lower Left: Sensory-evoked potentials were elicited by a random sequence of somatosensory, auditory, and visual stimuli. Auditory stimuli were brief tones produced by a speaker; visual stimuli were brief flashes produced by an LED; somatosensory stimuli were applied to the index finger of the left hand by a vibrator. Lower Right: Cognitive ERPs were elicited by the classical visual oddball paradigm with the red squares as the target stimuli and white as the nontarget stimuli on the screen.

Table 1
Data preprocessing sheet.

Software	Matlab 2018b & Letswave7 (Letswave.cn)
Band-pass filtering	Butterworth filter, 0.01–200 Hz, 4th order, 24 dB/octave, zero-phase
Notch filtering	Butterworth filter, 49–51 Hz, 4th order, 24 dB/octave, zero-phase
Channel interpolation	Bad channels were identified manually and interpolated with the mean value of the three surrounding channels.
Re-reference	Re-reference to the mean value of TP9 and TP10
Artifacts removal by ICA	Eye movement related ICA components were identified by visual inspection of their scalp topographies, time courses, and spectra.
Band-pass filtering	Butterworth filter, 0.1–30 Hz, 4th order, 24 dB/octave, zero-phase
Segmentation	Segmentation from −0.5 to 1.0 s relative to the stimulus onset
Averaging	Average all trials for each subject
Baseline correction	Remove the DC offset (based on −0.5 to 0 s pre-stimulus)

For data collection, EEG signals were recorded via a multichannel EEG system (64 Channel, Easycap) and an EEG Amplifier (BrainAmp, Brain Products GmbH, Germany). The signals were recorded at a sampling rate of 1000 Hz by 64 electrodes, placed in the standard 10–20 positions. FCz was set to be the reference. Before data acquisition, the contact impedance between the EEG electrodes and the cortex was calibrated to be lower than 20 kΩ to ensure the quality of EEG signals during the experiments.

2.1.4. EEG preprocessing

For EEG preprocessing (Pernet et al., 2020), the sequence of steps, specific parameters for each step in preprocessing pipeline are shown in Table 1. After preprocessing, grand average ERP waveforms were computed for each participant and stimulus type (visual, auditory, somatosensory, and target stimuli of the visual oddball paradigm). All EEG pre-processing steps were carried out by Letswave7 (Huang, 2019) and Matlab.

2.2. Reliability analysis

2.2.1. Peak-based analysis and pointwise analysis

As the peak of each ERP component indicates the time point with a larger signal-to-noise ratio in the surrounding samples, peak amplitude is commonly used as a representative feature in ERP analysis. In this research, the most significant positive and negative peaks were de-

tected by manually searching for the local maximum/minimum value in their corresponding time intervals for each subject. The mean amplitude around the peaks was not considered in this research because it is not fair to compare the reliability of pointwise analysis with the reliability of the mean amplitude, which is the average of multiple points.

Pointwise analysis was also used to examine the reliability of the ERP. More specifically, the ERP amplitude at each time point and each channel was taken as the variable for measuring the individual difference. Unlike the peak-based analysis, pointwise analysis is a fully data-driven method that is performed along with the temporal and spatial domain in a point-by-point way.

2.2.2. Metric of reliability: Intraclass correlation coefficient (ICC)

ICC is a commonly used metric for reliability analysis. In this study, the reliability was measured by using ICC(A, 1) of case 2A (McGraw and Wong, 1996) to represent the absolute agreement between repeated measurements for both the peak-based and pointwise analyses for both the peak-based and pointwise analyses. The subject-by-experiment matrix was modeled by a two-way ANOVA with random subject effects (row effects), random session effects (column effects), and residual effects, as shown in Eq. (1), and ICC(A, 1) is calculated as Eq. (2).

$$x_{ij} = \mu + r_i + c_j + e_{ij}, \quad (1)$$

$$ICC(A, 1) = \frac{\sigma_r^2}{\sigma_r^2 + \sigma_c^2 + \sigma_e^2}, \quad (2)$$

in which, $i = 1, \dots, n$ ($n = 82$) is used as the subscript for subjects; $j = 1, \dots, k$ ($k = 2$) is the subscript for multiple observations, which is the sessions in this work; x_{ij} denotes the observation in the j -th session from subject i ; μ is the population mean for all observations; $r_i \sim N(0, \sigma_r)$ represents the row effects for subject i ; $c_j \sim N(0, \sigma_c)$ represents the column effects for session j ; $e_{ij} \sim N(0, \sigma_e)$ represents the residual effects as the error terms. According to ICC(A, 1), the reliability was defined in Eq. (2) as the proportion of the between-subject variation σ_r^2 over the total variation $\sigma_r^2 + \sigma_c^2 + \sigma_e^2$. In Eq. (1), we assume r_i , c_j , e_{ij} are independent random variables with normal distribution. The mean value and confidence interval of ICC(A, 1) for both peak-based and pointwise analyses were obtained 1600 times by bootstrap, which involved choosing random samples with replacement from a dataset and analyzing each sample in the same way. Two sample t -test was applied to compare the result between peak-based and pointwise analyses. Based on Eq. (1) and Eq. (2), the variance in an ERP measure comes from three parts, which are $Var(Trait) = \sigma_r^2$, $Var(State) = \sigma_c^2$, and $Var(Noise) = \sigma_e^2$ (Segalowitz and Barnes, 1993). Partitioning variance into these three parts, temporal and spatial variation of the $Var(Trait)$, $Var(State)$, and $Var(Noise)$ was analyzed in the spatial and temporal domains of the ERPs in a pointwise way.

2.2.3. Statistical analysis

By comparing group effects and individual reliability, correlation analysis was performed between the reliability and each group- and individual-level measure. Taking AEP as an example, a one-sample t -test was performed point wisely at the post-stimulus time points from the ERP signal of 82 subjects against zero at electrode Cz. The time points, which were significantly different from zeros (p -value $< 0.05/1000$ by Bonferroni correction; 1000 was the number of post-stimulus time points) were selected to reduce the influences of noisy background activity. At these selected time points, two group-level measures and one individual-level measure were extracted. Two group-level measures were (1) **abs(t-value)** calculated by the absolute value of the t -value and (2) **Hilbert envelope** calculated by the grand value of the ERP envelope via the Hilbert Transform. The individual-level measure was (3) **between-subject variance** estimated by the standard deviation across the 82 subjects. Then the linear trends were removed from the time series of each measure and reliability to avoid spurious correlation. The associations between different measures and reliabilities were quantified by Spearman's rank correlation coefficient, which is more robust to the non-linearity of changes and outliers than Pearson's correlation. For SEP, VEP, and P300, the same procedures were applied at electrodes Cz, Oz, and Pz, respectively, to explore the consistency between group effects and individual reliability because those electrodes showed the strongest group-level response.

2.3. Model simulation

For the given real EEG data, we can calculate its reliability, but the underlying factors affecting reliability are fixed and unknown, which limits the further study of reliability. Hence, a simulation model is needed to investigate the underlying mechanism behind the observations from real data. As a supplement to the real EEG data analysis, a dynamic model simulation allows us to further understand the internal mechanism of the brain. As rhythmic oscillations are the basic characteristics of an EEG signal and evoked changes of an EEG signal could be ascribed to transients that arise as the system's trajectory returns to its attractor (David et al., 2005), in this work, a second-order linear model was applied in this study for its simplicity to explain what we observed in real ERP data (i.e., the inconsistency between group effects and individual reliability).

$$x'(t) = Ax(t) + C * \mu(t) + e(t), \quad (3)$$

in which $A = \begin{bmatrix} c & d \\ -d & c \end{bmatrix}$ is the state-transition matrix with the corresponding eigenvalues $c \pm di$, $c < 0$ to ensure that the real part of the

eigenvalue of A is negative which decides the decaying rate, d is the image part of the conjugate complex eigenvalues, which controls the natural oscillation frequency of this autonomous system. In this work, the elements of A were selected empirically as $c = -10$ and $d = 50$ to mimic the response of AEP at Cz. The influence of each parameter in A on model behavior was illustrated in the supplementary material (Fig. S6). The input strength, C , is formulated as $C_{sub} + C_{trial}$, where C_{sub} is a random variable representing the input strength for a given subject conformed to a Gaussian distribution (μ_{sub} , σ_{sub}^2), and C_{trial} is a random variable representing the input strength for a given trial conformed to a Gaussian distribution (μ_{trial} , σ_{trial}^2). According to Jansen and Rit's neural mass model (Jansen and Rit, 1995), the input of the system was simulated by using Eq. (4):

$$u(t) = \begin{cases} ate^{-bt} & t \geq jitter_{sub} \\ 0 & t < jitter_{sub} \end{cases} \quad (4)$$

in which $jitter_{sub}$ is a rounded random variable with a uniform distribution $[-\tau_{sub}, \tau_{sub}]$ relative to the onset time $t = 0$, and $e(t)$ is the pink Gaussian noise representing the input of the background EEG activity in the simulation. The core setting of this model was the additive term $C_{sub} + C_{trial}$, which coupled the input strength with the subject-level and the trial-level, thus allowing both $Var(Trait)$ and $Var(Noise)$ to co-vary with the signal amplitude. Considering the neglectable proportion of $Var(State)$ in real data results, there is no difference in simulation between sessions. To ensure consistency with real data, we set subject number $n = 82$ and session number $k = 2$ in the simulation. For each session, there are 60 trials. To mimic the real data preprocessing procedure, baseline correction was also applied to simulated ERP. With a sampling rate of 1000 Hz, there were 1500 time points for each trial, from -0.5 to 1 s.

In this study, two major parameters of this model potentially influencing the test-retest reliability were investigated: (1) inter-subject variability, τ_{sub} , for the latency jitter, $jitter_{sub}$, and (2) inter-trial variability, σ_{trial} , for the input strength, C_{trial} . Considering the oscillation of the ERP response as the trajectory in the 2-dimensional phase portrait in Fig. 2C, the observed ERP response is the projection of this trajectory on the axis of x_1 . Hence, the peaks, troughs, and zero crossings have no special meaning, but some specific phases when the trajectory rotates along with the origin. The value of σ_{trial} affect the magnitude of the ERP trajectory. Hence, σ_{trial} determine the disturbance normal to the trajectory of the ERP response. While the value of τ_{sub} affect the time of trajectory of the ERP response. Hence, τ_{sub} determine the disturbance tangent to the trajectory of the ERP response. These two factors, τ_{sub} and σ_{trial} , were selected to investigate the test-retest reliability in the simulation because they provide disturbances in two directions orthogonal to each other. The different influence in different phases of the ERP response was expected for these two factors. Further, σ_{trial} is a trial-level factor, τ_{sub} is a subject-level factor, and the change in $Var(Trait)$ and $Var(Noise)$ could be investigated in the simulation. The simulation code is available online (<https://osf.io/v59qu>).

3. Results

3.1. Reliability of real data

3.1.1. Reliability for multisensory and cognitive ERPs

The grand average waveform of AEP at channel Cz, SEP at channel Cz, VEP at channel Oz, and P300 at channel Pz are shown in Fig. 3, where red and blue curves shaded by the standard deviation denoted the signals of the two sessions. The representative ERP peaks, including N1 at 90 ms and P2 at 180 ms for AEP, N2 at 150 ms and P2 at 245 ms for SEP, N1 at 64 ms, P2 at 185 ms for VEP, and P3 at 345 ms of P300, were selected for peak-based analysis. The negative and positive peaks are indicated by light gray and dark gray lines respectively and the gray shaded interval indicated the interval between the two peaks. For the pointwise analysis, the thick black lines indicated the maximal relia-

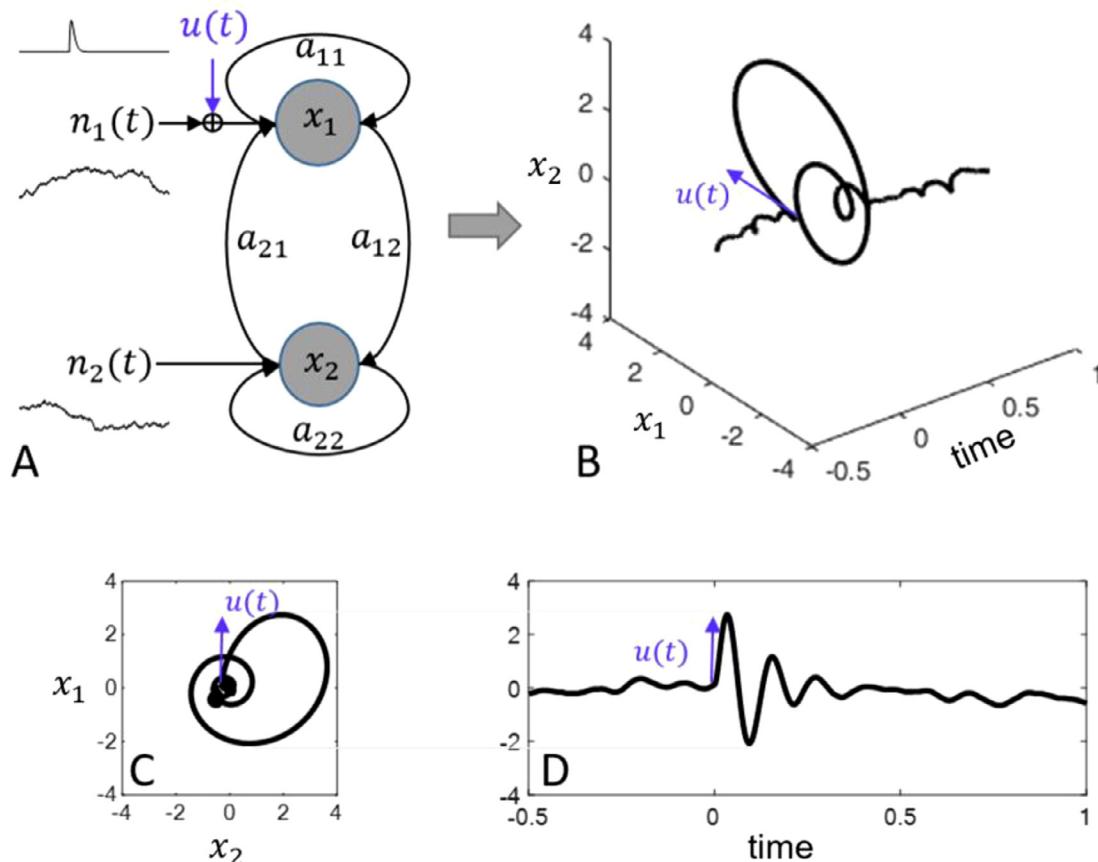


Fig. 2. The ERP simulation, which was generated by a second-order dynamic system model (3). (A) the framework of the dynamic model in Eq. (3); (B) the evolution of x_1 and x_2 over time, with its two-dimensional projection of the phase portrait in (C) and the evolution of x_2 over time in (D).

bility along with the ERP time courses. Since the subjects were more familiar with the experimental environment in the second session, the amplitudes of the ERP peaks were reduced. Importantly, the most reliable time point in the ERP did not correspond to the ERP peak. For P300 shown in Fig. 3D, the maximal reliability time point in the pointwise analysis appeared at 223 ms (thick black line), with a reliability of 0.61. This was much earlier than the well-known P3 component in the peak-based analysis, with the reliability of 0.44.

From each violin plot, the high reliability depended on (1) the consistency of individual ranks between two sessions and (2) the large between-subject variances. Taking the pointwise ICC results in AEP for example, the amplitude of ERP showed comparable inter-subject variance at time point 90 ms, 133 ms and 180 ms. But at 133 ms, the less interleaving between session 1 and 2 leads to a higher value of ICC (0.76) as compared with the other two time points. Taking the pointwise ICC results in VEP for another example, the inter-subject variances at 64 ms was relatively lower than time points at 183 ms and 185 ms, thus its reliability was much lower.

The comparison between the reliability results of the peak-based analysis and the pointwise analysis at corresponding time points were shown in Table 2, in which the peak amplitude was significantly ($p < 10^{-4}$) less reliable than the corresponding pointwise amplitudes at the latency of the grand average for all four types of ERPs.

3.1.2. Spatiotemporal evaluation of reliability: a case study of AEP

Next, AEP was used to further investigate the consistency between group effects and reliability with different exploratory analyses (results for other types of ERPs are provided in the supplementary material (Fig. S(1–4))). As illustrated in Fig. 4A, the t -value of significant regions (p -value $< 0.05/1000/64$ with Bonferroni correction, where 1000 is the

Table 2

Comparisons between the reliability of peak amplitudes and corresponding pointwise amplitude for AEP, SEP, VEP, and P300.

ERPs	Measurements	Reliability		p -value
		Mean	Conf	
AEP(Cz)	N1	0.50	[0.31,0.68]	$< 10^{-12}^{***}$
	90ms	0.54	[0.34,0.70]	
	P2	0.56	[0.39,0.68]	$2.8 \times 10^{-8}^{**}$
	180ms	0.58	[0.43,0.70]	
	133ms	0.76	[0.68,0.85]	
	110ms	0.70	[0.60,0.83]	
SEP(Cz)	N2	0.54	[0.38,0.68]	$< 10^{-12}^{***}$
	150ms	0.57	[0.43,0.71]	
	P2	0.64	[0.52,0.78]	$2.2 \times 10^{-8}^{**}$
	245ms	0.66	[0.54,0.79]	
	64ms	0.48	[0.34,0.62]	$< 10^{-12}^{***}$
VEP(Oz)	P2	0.54	[0.33,0.69]	
	185ms	0.73	[0.61,0.84]	
	183ms	0.74	[0.62,0.84]	
	185ms	0.74	[0.62,0.84]	
P300(Pz)	P3	0.44	[0.24,0.63]	$< 10^{-12}^{***}$
	345ms	0.50	[0.31,0.66]	
	223ms	0.61	[0.48,0.73]	

number of post-stimulus time points, and 64 is the number of channels) were presented in the shaded region, which was consistent with the amplitude of the grand average waveform. Also, the post-stimulus AEP response behaved as a process of attenuating oscillations and finally approached the baseline. In contrast, the reliability of AEP after the stimulation shown in Fig. 4B increased greatly at the beginning of stimulation, lasting for a certain period, and then slowly returned to 0.

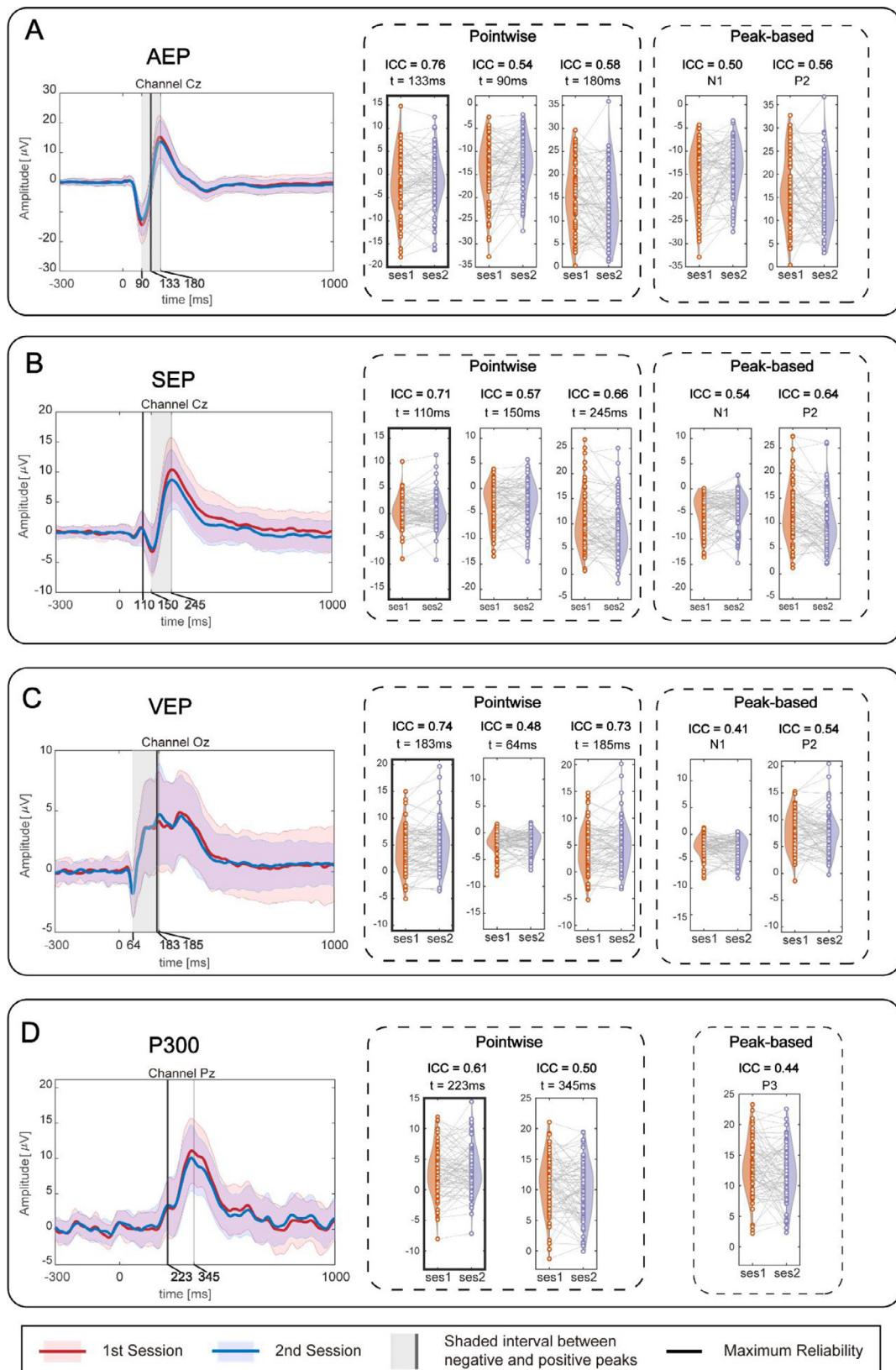


Fig. 3. Grand average waveform and the test-retest reliability result for AEP, SEP, VEP, and P300. The left part is the grand average waveform for the four types of ERP, shaded by the standard deviation in both 1st session (in red) and 2nd session (in blue). The gray shaded interval indicates the interval between the negative and positive peaks. The thick black line indicates the time point for the maximum ICC value. On the right part, the violin plot shows the amplitude distribution and change of amplitude for each subject between two sessions for both pointwise comparison and peak-based comparison.

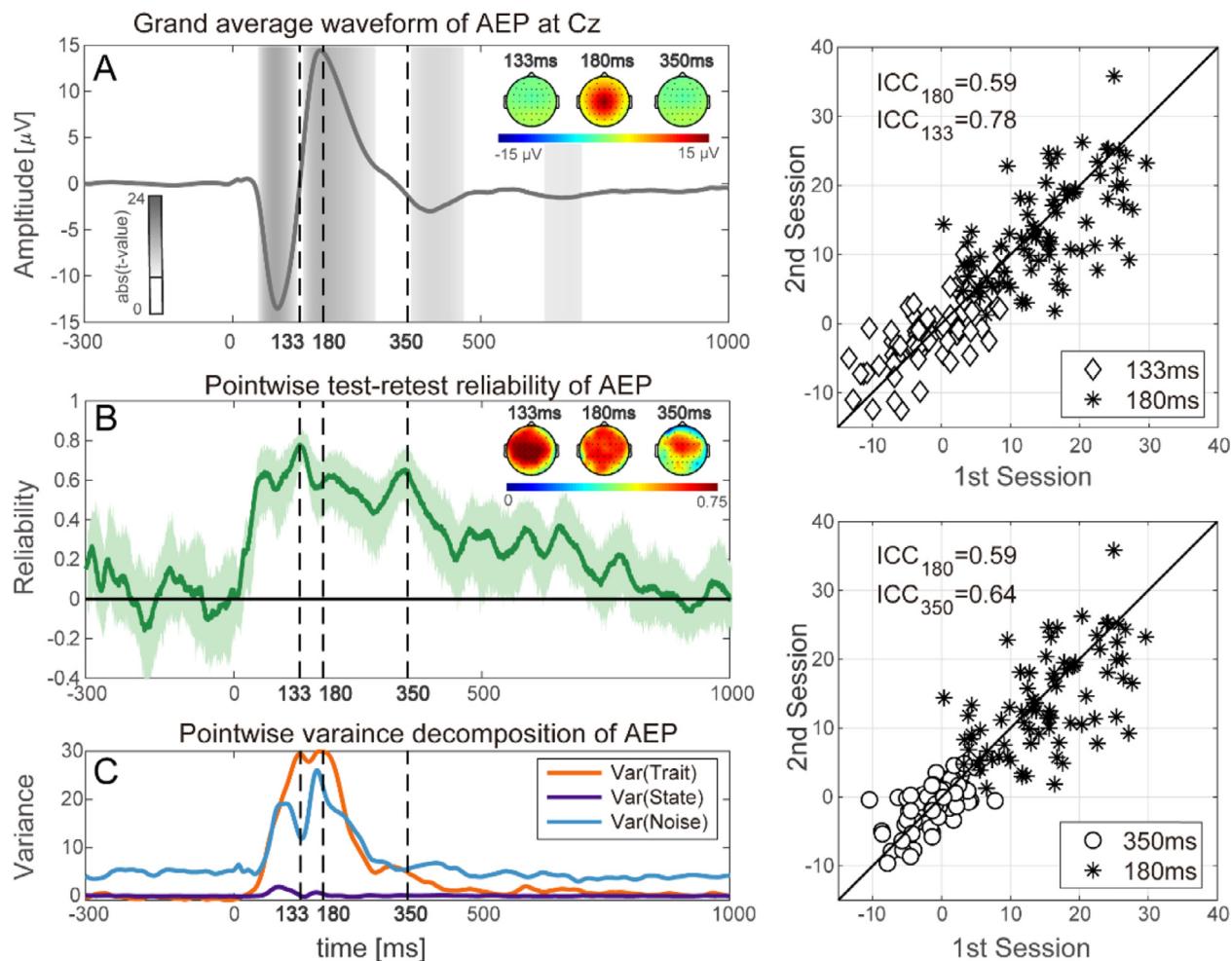


Fig. 4. The test-retest reliability analysis for AEP at electrode Cz. (A) The grand average waveform of AEP. (B) Pointwise test-retest reliability analysis considering the entire shape of the AEP time-course calculated by ICC(A,1). (C) The variance in the observation matrix with the size of subject \times experiment was decomposed into three parts: $Var(Trait)$, $Var(State)$, and $Var(Noise)$ by a two-way random effects model along with the AEP time-course. Scatter plots of 82 subjects' amplitudes of AEP at Cz in the two experiments were compared between 180 and 133 ms in (D) and between 180 and 350 ms in (E).

Hence, the reliability of AEP along the ICC temporal profiles was not correlated with the amplitude of AEP. The maximal reliability of 0.78 appeared at the time point 133 ms, with the mean amplitude of AEP close to 0, which did not correspond to N1 at 90 ms or P2 at 180 ms with the minimal/maximal amplitude. The topographies of the grand average amplitude and the reliability at 90, 133, and 180 ms are illustrated in Figs. 4A and 4B.

Pointwise variance decomposition results based on a two-way ANOVA are shown in Fig. 4C. The magnitude of $Var(Trait)$ was close to 0 before the stimulation. At the beginning of stimulation, the magnitude of $Var(Trait)$ reached a peak at 180 ms and then returned to 0. The local maximum of $Var(Trait)$ did not correspond to the peak of AEP. The magnitude of $Var(Noise)$ showed similar trends as $Var(Trait)$, but the baseline was not 0. During the first 400 ms after stimulation, there was a certain correspondence between the waveform of AEP and $Var(Noise)$. The peak of AEP corresponded to the local maximum of $Var(Noise)$, while the zero-crossing point of AEP corresponded to the local minimum of $Var(Noise)$. Compared with $Var(Noise)$ and $Var(Trait)$, the magnitude of $Var(State)$ was too small and had little impact on reliability. Hence, the reliability was mainly determined by the ratio of $Var(Trait)$ to $Var(Noise)$. Next, the time points of 133, 180, and 350 ms were selected for the comparison, in which 180 ms corresponded to the peak of the grand average of AEP, while 133 and 350 ms corresponded to the local maximum of the reliability. It should be noted that, due to the

insufficient session number ($n = 2$), the estimation of $Var(State)$ would possibly be negative, which was deeply elaborated in the supplementary material (Fig. S5).

Fig. 4D shows the comparison of the scatter plots of 82 subjects' amplitude of AEP between time points 133 ms (diamonds) and 180 ms (asterisks). As $Var(State)$ was close to 0, $Var(Trait)$ could be measured as the variance along the black diagonal line, and $Var(Noise)$ could be measured as the variance perpendicular to the black diagonal line. As shown in Fig. 4D, the mean amplitude of AEP at 133 ms (mean value of the diamonds) was much smaller than that at 180 ms (mean value of the asterisks), but the $Var(Trait)$ values at the two different time points were similar. Hence, the reliability at 133 ms was larger than that at 180 ms because of the smaller $Var(Noise)$ at 133 ms. Fig. 4E shows a different situation compared with that shown in Fig. 4D. The reliability at 180 ms (asterisks) and 350 ms (circles) were similar, but both $Var(Trait)$ and $Var(Noise)$ at 350 ms were smaller than that at 180 ms.

3.1.3. Statistical results

Spatiotemporal dissociation between group effects and individual reliability was revealed in Fig. 3 and Fig. 4. These findings went against our expectations, given the fact that extracting peak-based measures using group-level prior information was the most common approach in reliability analysis. Hence, Spearman's rank correlation analysis was further performed on AEP, SEP, VEP, and P300 to analyze the statistical

Table 3

Associations between group-level measures ($\text{abs}(t\text{-value})$, Hilbert envelope), individual-level measure (between-subject variance), and reliability.

ERPs	abs(t -value)		Hilbert envelope		between-subject variance	
	Spearman's ρ	p-value	Spearman's ρ	p-value	Spearman's ρ	p-value
AEP (Cz)	-0.19	3.04×10^{-5}	0.27	$< 10^{-12}$	0.36	$< 10^{-12}$
SEP (Cz)	0.38	$< 10^{-12}$	0.51	$< 10^{-12}$	0.71	$< 10^{-12}$
VEP (Oz)	0.17	0.002	0.12	0.017	0.75	$< 10^{-12}$
P300 (Pz)	0.54	$< 10^{-12}$	0.74	$< 10^{-12}$	0.84	$< 10^{-12}$

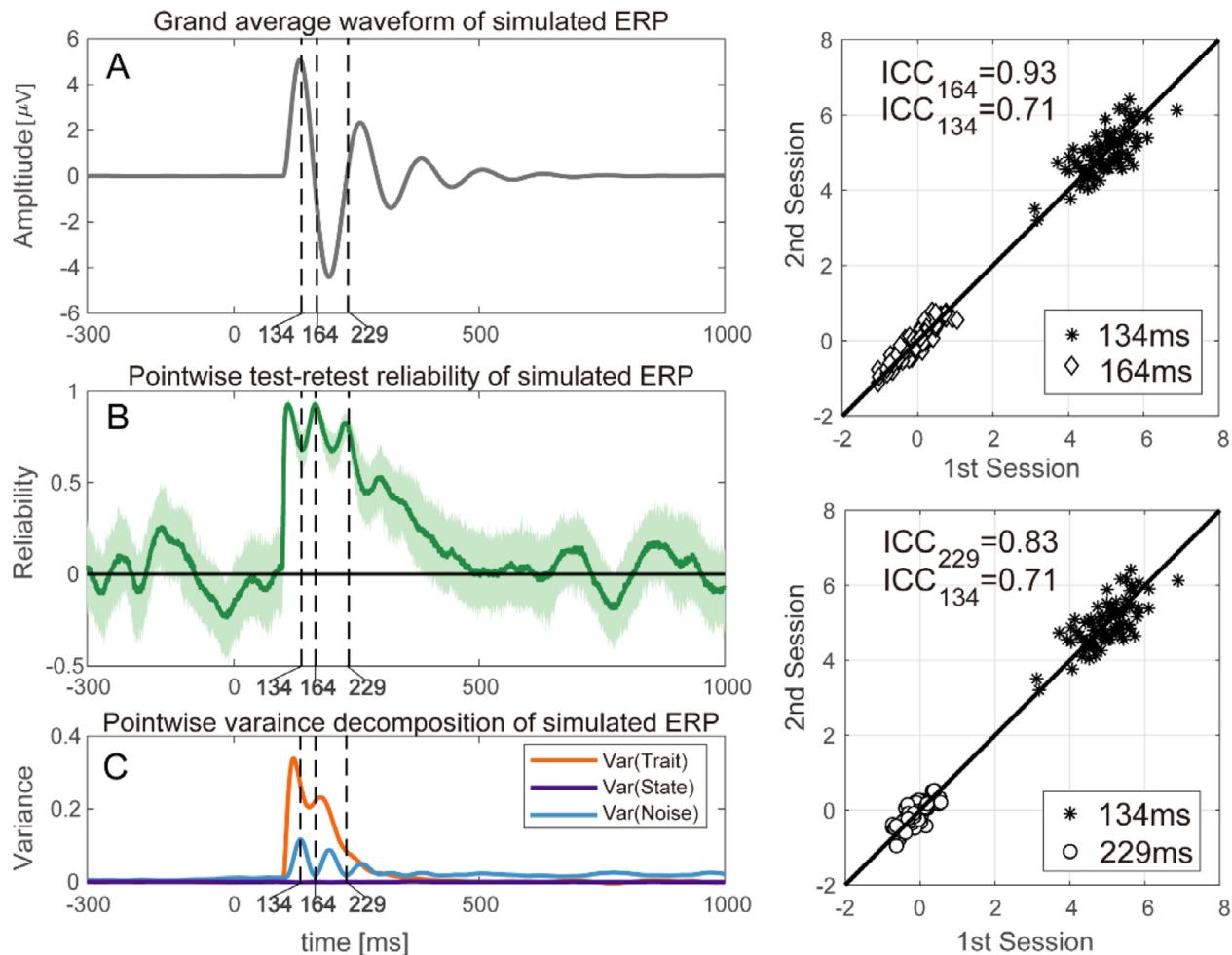


Fig. 5. The test-retest reliability analysis for simulated ERP. (A) The grand average waveform of simulated ERP for a given set of system parameters. (B) Pointwise test-retest reliability analysis along with the simulated ERP time course. (C) The variance of observation matrix with the size of the subject by experiment was decomposed into three parts: $\text{Var}(\text{Trait})$, $\text{Var}(\text{State})$, and $\text{Var}(\text{Noise})$ by a two-way random effects model along with the simulated ERP time-course. Scatter plots of 82 subjects' amplitudes of AEP at electrode Cz in two experiments were compared between 134 and 164 ms in (D) and between 134 and 229 ms in (E).

relationships of reliability with the group-level measures ($\text{abs}(t\text{-value})$ and Hilbert envelope) and individual-level measure (between-subject variance). For group-level measures, it was observed that the Hilbert envelope showed a larger correlation coefficient with reliability than the $\text{abs}(t\text{-value})$ except for VEP. This result suggested an oscillation perspective on ERP reliability should be considered. For individual-level measure, Spearman's ρ between between-subject variance and reliability was greatly improved (Table 3).

3.2. Reliability of simulated data

3.2.1. Simulation results

To further understand the internal factor influencing the reliability in ERP analysis, a dynamic model in Eq. (3) was used for the simulation. The simulation results in Fig. 5 were consistent with the results

from real ERP data in Fig. 4. Specifically, the grand average waveform of the simulated ERP is shown in Fig. 5A, with a peak of 134 ms and subsequent zero crossings appearing at 164 and 229 ms. The reliability curve across time is shown in Fig. 5B, and the corresponding variance decomposition is shown in Fig. 5C. In the simulation, the correlation between $\text{Var}(\text{Noise})$ and the amplitude of the ERP was more obvious. $\text{Var}(\text{State})$ was close to 0 because the systematic differences between the two sessions were not considered in this simulation. Hence, the reliability at the peak latency was the local minimum, and the reliability at the zero-crossing point was the local maximum. Similarly, the scatter plots in Figs. 5D and 5E show that the larger amplitude of the ERP may not necessarily lead to greater reliability, which was determined by the ratio of $\text{Var}(\text{Trait})$ to $\text{Var}(\text{Noise})$.

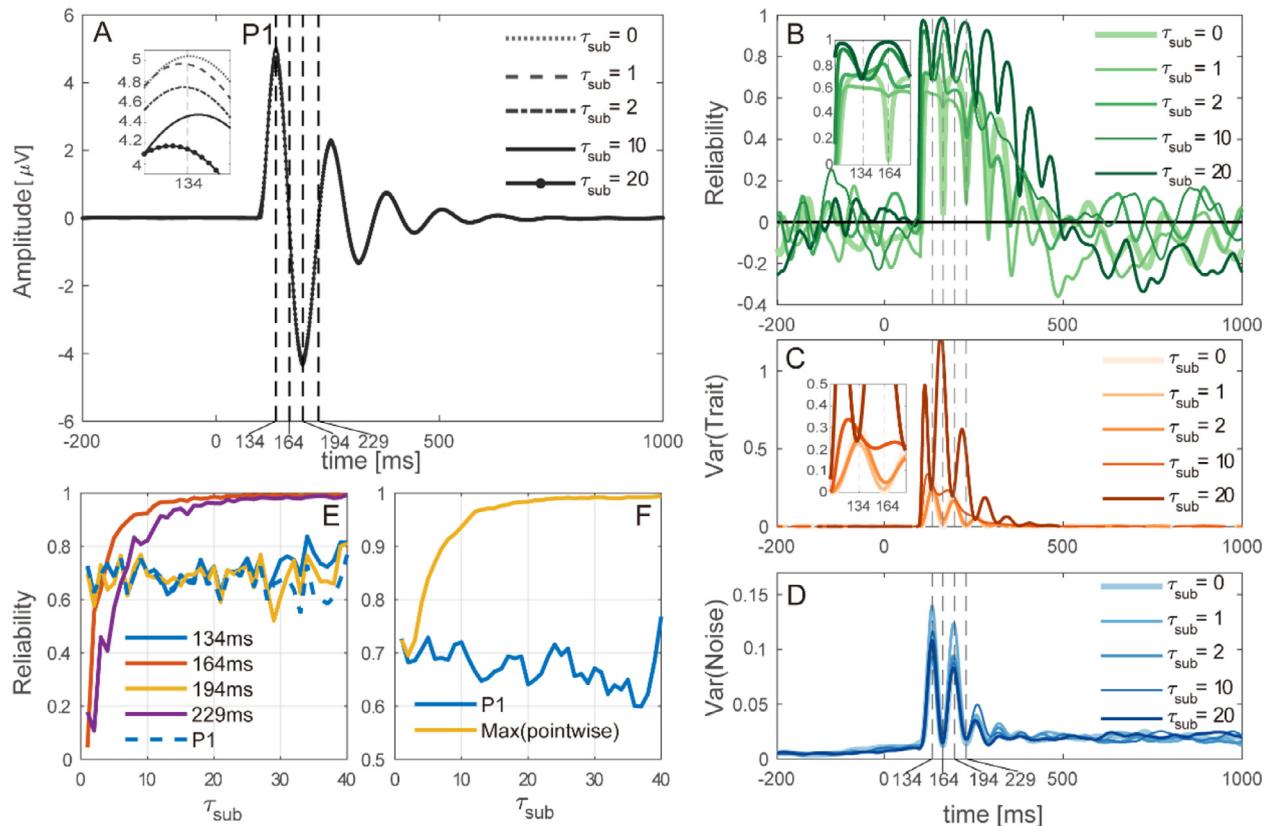


Fig. 6. The influence of increasing the variability of inter-subject latency jitter of the dynamic system at the subject-level on (A) Grand average waveform of simulated ERP. (B) Pointwise test-retest reliability along the time-course of simulated ERP. (C) $Var(\text{Trait})$ along the time-course of simulated ERP. (D) $Var(\text{Noise})$ along the time-course of simulated ERP. (E) Comparisons between peak-based reliability and pointwise reliability at group-level peak latencies. (F) Comparisons between the maximum value of pointwise reliability and peak-based reliability.

3.2.2. The influence of the variability of jitter: τ_{sub}

As a tangential disturbance in the phase portrait of Eq. (3) shown in Fig. 2, an increase in τ_{sub} did not make a large difference in the waveform of the grand average ERP in the simulation, but it made the peak of P1 and N2 smoother. As τ_{sub} increased from 0 to 20, the amplitude of the peak P1 reduced slightly, as shown in Fig. 6A. As illustrated in Fig. 6B, the reliability of the peaks at 134 and 194 ms remained around 0.7, while the reliability of the zero-crossing points at 164 and 229 ms increased greatly. As inter-subject latency jitter increased, the ICC values at 134 and 194 ms, which corresponded to the peaks of the grand average waveform, gradually shifted from the peaks of the ICC temporal profiles to their local minimum. The ICC values at 164 and 229 ms, which corresponded to the zero-crossing point, behaved conversely. The corresponding variance decomposition is shown with different values of τ_{sub} in Fig. 6(D–E). $Var(\text{Trait})$ at the zero-crossing point (164 and 229 ms) of the grand average waveform increased as inter-subject latency jitter increased, while $Var(\text{Noise})$ fluctuated randomly. In comparison with the reliability of the peak amplitude, there was a greater difference between the maximum values of the ICC temporal profiles and the reliability of the peak amplitude.

3.2.3. The influence of the variability of input power: σ_{trial}

For normal perturbation in the phase portrait of Eq. (3) in Fig. 2, it is shown in Fig. 6B that the overall magnitude of the ICC temporal profiles dropped because of increasing inter-trial variability in the dynamic systems' input, while the reliability in the response amplitude at 134 and 194 ms, which corresponded to the peak latencies of the grand average waveform, dropped more quickly compared with the reliability of the response amplitude at 164 and 229 ms, exhibiting an unbalanced influence. The above observations were further investigated by

pointwise variance decomposition, in which the within-subject variation ($Var(\text{Noise})$) increased systematically in proportion to the signal amplitude of the grand average waveform with increasing inter-trial variability, while the between-subject variation fluctuated randomly, thus explaining why the reliability of the signal with larger amplitude decreased more. Interestingly, it can be noted from Fig. 6F that there was a larger difference between the maximum values of the ICC temporal profiles and the reliability of the peak amplitude as the inter-trial variability increased.

4. Discussion

The purpose of this study was to investigate the relationships between group effects and individual reliability across different types of ERPs. By performing pointwise reliability analysis and rigorous simulation, we found inconsistency between individual reliability and group effects and provided potential explanations from the perspective of oscillations of ERP. The findings have implications for a series of questions that are of theoretical and practical relevance for ERP researchers, which will be discussed sequentially.

4.1. Peak-based analysis versus pointwise analysis

By briefly reviewing the ERP reliability research in the past decade in Table 4, we think it is necessary to re-emphasize that we should not restrict analysis in narrow time windows around peaks, especially for research about the individual difference. Until now, peak-related feature extraction (i.e., peak amplitude, area under the curve, mean amplitude) has been a dominant approach for examining the reliability of ERPs (Huffmeijer et al., 2014; Munsters et al., 2019; Devos et al.,

Table 4
Literature review on ERP reliability from 2011 to 2021.

ERP reliability research	
Gaspar et al., 2011	Grand average ERPs can be misleading because it does NOT reflect individual dynamics. Not only around the peak, but also entire temporal windows are reliable.
Cassidy et al., 2012	Reliability of peak and latency for a selected range of ERP components were evaluated.
Leue et al., 2013	Intra-individual N2 variability incorporated systematic variance.
Huffmeijer et al., 2014	ERP amplitudes generally showed adequate to excellent test-retest reliability.
Ip et al., 2018	Averaging across several electrodes or trials improved the reliability of P3 amplitude.
Munsters et al., 2019	ERP measures exhibited more variation and are less stable compared to continuous EEG.
Our research	The face-sensitive ERP components (i.e. N290, P400, and Nc) in infants show adequate test-retest reliability
	<ul style="list-style-type: none"> • The peak-based analysis may not be sufficiently reliable to capture the individual difference. • A perspective of neural oscillations is more peak-based analysis to explain the inconsistency between group effects and individual reliability. • A simulation model is applied to investigate underlying factors of modulating the consistency between the group effect and individual reliability.

2020). For the peak-based approach, researchers have found that the reliability of ERPs is influenced by the number of trials, channel selection, and various preprocessing strategies (Huffmeijer et al., 2014; Leue et al., 2013). The basic hypothesis behind the peak-based analysis is that the peak of the ERP indicates a higher signal-to-noise ratio, which produces results with a higher confidence level because of the relatively small interference from background EEG noise, yet this concept of signal-to-noise ratio may not generalize to the research area interested in individual difference, in which between-subject variance is treated as the signal, within-subject variance is treated as noise, as mentioned by (Brandmaier et al., 2018). Another limitation of the peak-based analysis is that the latency and amplitude of ERP peaks, as well as the entire ERP shapes, are physiologically meaningful and important (Gaspar et al., 2011).

Further, we also observed the phenomenon that ERP shapes in different sessions for the same subject were quite similar, while ERP shapes for different subjects were largely different, which may reflect that the information processing underlying ERP is unique to each subject. We believe future ERP research should not restrict its analysis in narrow time windows around peaks, but develop analysis techniques to characterize the ERP shape, which is useful in translational neuroscience.

4.2. Inconsistency between group effects and individual reliability

Stronger group effects do not guarantee higher individual reliability. In reliability analysis, the group effects are commonly used as prior information (Plichta et al., 2012; Aron et al., 2006; Fliessbach et al., 2010), which assumes that experimental manipulation eliciting greater activation at the group level should also show reliable between-subject variation. This conventional approach has been questioned in recent years, especially in the fMRI community (Fröhner et al., 2019; Infantilino et al., 2018; Yarkoni and Braver, 2010; Li et al., 2019). In line with these studies, our results also revealed inconsistency between group effects and individual reliability in ERP analysis. More specifically, concerning the temporal domain discrepancies illustrated in Fig. 3, the most reliable points in the four types of ERPs (Cz for AEP, SEP, Oz for VEP, and Pz for P300) did not all correspond to maximum or minimum points of group-level activations. For AEP, the most reliable point appeared at the zero-crossing point of ERP. The spatial domain discrepancies are illustrated in Fig. 4 for AEP, in which we did not find the topography of the AEP response corresponding to the topography of the reliability at 133 and 350 ms. Further analysis, presented in Table 1, indicated that, as an individual-level measure, the between-subject variance showed a higher correlation coefficient than the group-level measure ($\text{abs}(t\text{-value})$) across all four types of ERPs. All these evidences suggest that it is not advisable to select peak-related features at the electrode showing the strongest group effects without carefully examining their reliability.

Intuitively, the spatial-temporal distribution of group-level analyses and individual-level analyses should tend to converge. In other words, increased activation by experimental manipulation at the group level

should relate to individual-level analysis, given a large enough sample size and no confounding factors. However, few empirical pieces of evidence support this idea (Lee et al., 2006); more often, individual difference analyses simply fail to reveal any significant effects in regions that show a robust within-subject effect (Vetter et al., 2017; Raemaekers et al., 2007). In this research, the simulation results indicate that the consistency between group-level effects and individual reliability may be dynamically modulated by inter-subject latency jitter and inter-trial variability of dynamic system input, providing a dynamic view of the relationships between the two types of analysis in ERP analysis.

Spatiotemporal evaluation and decomposition of reliability are good for identifying the reproducible individual difference. In this research, it was found that the high signal-to-noise ratio assumption for the peak of ERP did not hold when considering individual difference research, which was also mentioned by Brandmaier et al. (2018). As illustrated in Fig. 4, the variance of the noise (blue curve in Fig. 4C) was highly correlated with the magnitude of the AEP response (absolute value of the black curve in Fig. 4A). Considering that the essence of EEG is neural oscillation, the peak in the ERP is just a certain phase (0 or π) during the oscillation. There is nothing more special about it compared to other phases. Hence, there is no reason why reliability analysis should be limited to peak-based features; pointwise analysis can bring us more comprehensive results. As compared with t-test or ANOVA, the pointwise analysis of test-retest reliability did not have the family-wise error rate problem, as we calculated the ICC values but did not judge whether there was a significant difference. Compared with peak-based analysis, the results from the pointwise analysis always had significantly higher ICC values at the time point of the peaks for all four types of ERP analysis in our investigation. In the test-retest reliability of AEP, SEP, VEP, and P300, the pointwise analysis consistently showed that the ICC value increased greatly after the stimulation, and after maintaining it for a while, decreased slowly to the baseline (Fig. 4). Hence, the peak of ERP may not relate to a higher ICC value. Even in AEP, the two local maximum points of the ICC value corresponded to the two zero-crossing points of the AEP. The findings suggest that reliability analysis restricted by the narrow time windows around the peaks is questionable. By performing pointwise analysis, dynamic changes in reliability in the spatial-temporal domain can be traced, given enough sample size, thus providing a new angle of ERP reliability analysis in a data-driven manner. Also, agreeing with the opinions of Gaspar et al. (2011), we believe that shape-based metrics rather than peak-based metrics may be more reliable for individual difference research.

To translate group effects into individual difference research, some issues must be reconsidered in ERP data analysis. ERP analysis focusing on individual difference often implicitly or explicitly uses prior information from group effects. For reliability analysis, electrodes showing the strongest stimulus-related activity by group-level analysis are often chosen for test-retest reliability analysis of ERPs (Gaspar et al., 2011). For constructing single-subject predictive models, it has been done in trans-

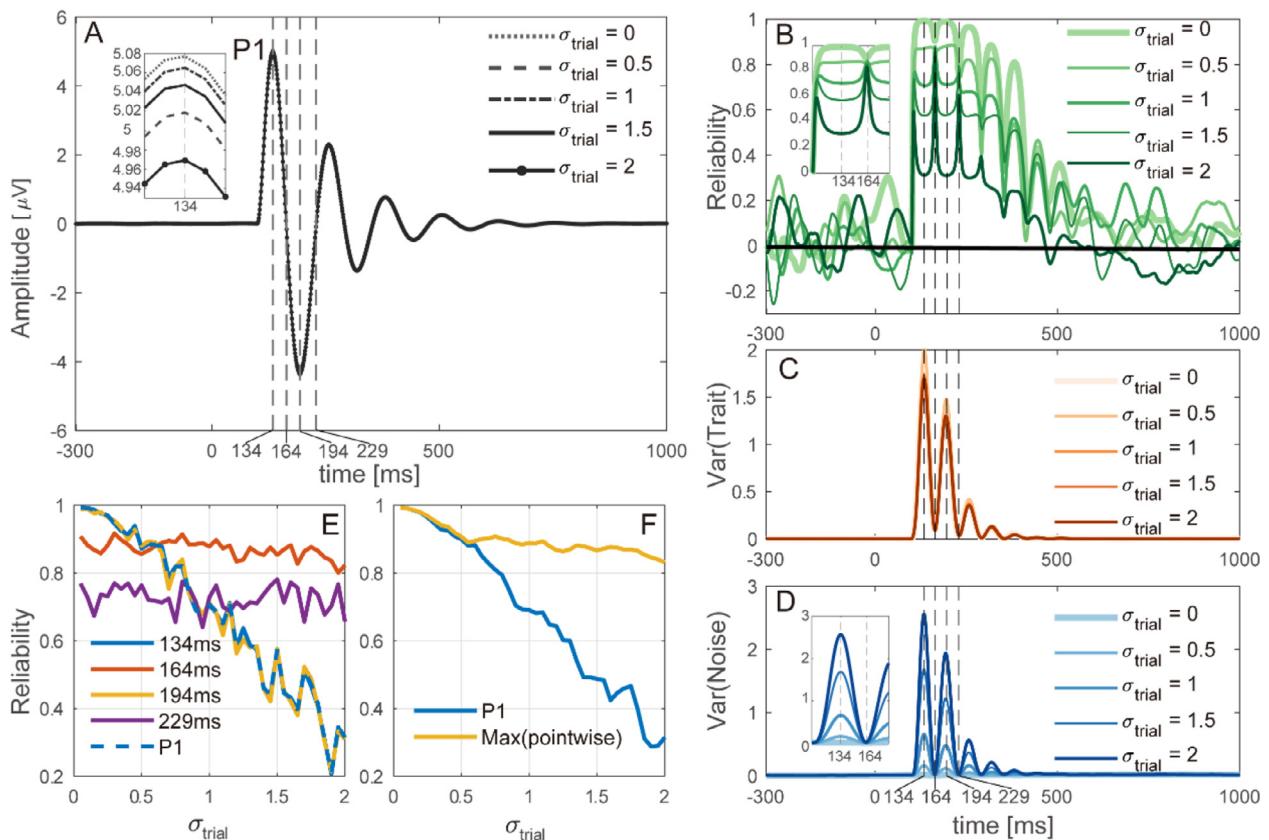


Fig. 7. The influence of increasing the variability of input power of the dynamic system at the trial-level on (A) Grand average waveform of simulated ERP. (B) Pointwise test-retest reliability along the time-course of simulated ERP. (C) $Var(Trait)$ along the time-course of simulated ERP. (D) $Var(Noise)$ along the time-course of simulated ERP. (E) Comparisons between peak-based reliability and pointwise reliability at group-level peak latencies. (F) Comparisons between the maximum value of pointwise reliability and peak-based reliability.

lating findings in group-level statistics of ERPs into a machine learning framework (Boshra et al., 2019). Hence, here we discuss some critical issues about ERP data analysis concerning the use of group effects in individual difference research.

- Tracing back to history, the original idea of ERPs was to index specific cognitive processes rather than distinguishing different individuals (i.e., the research interest was how brain activity responds to one condition versus the other). Individual differences were treated as measurement error that could not be explained by experimental manipulation, as t -test, omnibus ANOVA assumes. From this perspective, there is no reason to select regions based on strongest group effects and then feed them into the correlation or reliability analysis, except that this region also shows greater between-subject variation and smaller within-subject variation.
- For the data analysis pipeline of ERPs, it is very common to perform the subtraction operation (e.g. ERP difference waves) to minimize the impact of baseline individual differences. Such operations forcibly promote the activity of the baseline period at a constant rate across individuals, and the goal is to obtain a reliable experimental effect at the group level. This pervasive practice was inherited from research focusing on experimental effects, but few studies have noticed whether this approach is reasonable for individual difference analysis. Recently, the fMRI and psychology communities have argued that difference scores often exhibit a robust group-level effect but lower reliability (Infantilino et al., 2018; Onie and Most, 2017).

Considering statistical analysis, typically, ERP data are averaged within conditions and participants after preprocessing and then analyzed for the mean difference between conditions using paired t -test or repeated-measures ANOVA. This traditional approach implicitly as-

sumes that experimental manipulation yields uniform effects across all participants. The random variance of individual difference in effect sizes is not taken into account. By adopting linear mixed-effect models, in which random effects are used to capture individual variability as a form of random slopes or random intercepts, fixed effects are estimated by the grand mean across all participants. Such an approach has been adopted to simultaneously capture both group effects and individual difference (Frömer et al., 2018; Tibon and Levy, 2015).

4.3. A neural oscillation perspective on ERP reliability

Both peak and zero-crossing points of ERPs just represent different phases of one unified oscillation process. To further understand spatial-temporal inconsistency between group-level effects and individual reliability in ERP analysis, a dynamic model was applied for the simulation. The simulation model was simplified to be a second-order linear attractor with noise to simulate the EEG oscillation. From the perspective of dynamic system theory (Jansen and Rit, 1995; Yousofzadeh et al., 2015), peaks in the EPR are just an observation of EPR from one dimension of the computational models of neural processes. The phase portrait of our simulations (Fig. 2) provides a more comprehensive perspective, in which the peaks are just some special phases during the neural oscillation. Consider the oscillation of the EPR response as the trajectory in the 2-dimensional phase portrait as illustrated in Fig. 2C, the EPR response we observed is the projection of this trajectory on the axis of x_1 . Hence, the peaks, troughs, and zero crossings have no special meaning, but some specific phases when the trajectory rotates along with the origin. The value of σ_{trial} will affect the magnitude of the EPR trajectory. Hence σ_{trial} determined the disturbance normal to the trajectory of the EPR response. While the value of τ_{sub} will affect the time of trajec-

tory of the ERP response. Hence τ_{sub} determined the disturbance tangent to the trajectory of the ERP response. Owing to the different directions and different levels of the two factors, the simulation result showed that the changes of these two factors played very different roles in different phases of ERP, which can be illustrated by Fig. 6 and Fig. 7, especially at the peak and zero-crossing points of the ERP, with the changes of these two factors. In summary, With the similar wave form of the group-level ERP, the reliability would be determined by several factors. Measuring the peak-based features would not provide a comprehensive understanding about the oscillations in ERP.

Considering that stronger group effects do not guarantee higher individual reliability and the oscillation nature of ERP, the Hilbert transform was performed on trial-averaged data for each subject. The results in Table 3 suggest the Hilbert envelope is more consistent with reliability compared with the abs(t-value), which reflects the oscillation nature of ERP. The consistency between the grand averaged envelope of Hilbert transformed data and the reliability of ERP waveform for four kinds of ERPs without Bonferroni correction can be found in Fig. S7 in the supplementary material. For AEP, SEP, VEP, and P300, compared with the grand average ERP waveform, the grand average envelope of Hilbert transformed data is more consistent with individual reliability and shows a larger correlation coefficient, especially for AEP and SEP. These results further solidify the oscillation nature of ERP.

4.4. Limitations

Our research on reliability analyses had several limitations. First, higher reliability does not ensure higher validity. The fact that the response amplitude at some time points was more reliable than the peak amplitude may be explained by sacrificing validity. More specifically, each subject's response amplitude at a given time point may index different neurophysiological processes, leading to larger between-subject variance. Increasing reliability in this way is not desirable because the underlying process of this measure is different across subjects. However, we cannot verify this potential explanation without behavioral data. Second, the insufficient section number ($k = 2$) would lead to an inaccurate estimation of $Var(State)$, which make the negative value of $Var(State)$ possible in the practical calculation. Third, our analysis was restricted to univariate features; the relationship between group-level effects and individual reliability concerning multivariate analysis warrants further investigation in the future.

5. Conclusion

In summary, the purpose of this research was to investigate the consistency between group effects and individual reliability of ERPs. We performed spatiotemporal evaluation and decomposition of reliability in four different ERPs, and the findings indicate that the peak-based approach (i.e., selecting regions showing the strongest group-level response as individual difference variables) may be inappropriate for reliability analysis of ERPs. Without carefully examining reliability, this approach based on group-level prior information may fail to reliably capture individual differences, which is supported by spatiotemporal dissociation between group effects and individual reliability. The disadvantages of peak-based reliability analysis were illustrated by spatiotemporal evaluation and decomposition of reliability, statistical results, and the phase portrait in the simulation model. Further, the simulation results highlight the modulation role of inter-subject latency jitter and inter-trial variability in modulating the consistency between group-level effects and individual reliability. To conclude, all these results provide a new perspective beyond peak-based analysis in the ERP reliability studies. Furthermore, the findings deepen our understanding of ERP generation and the reliability of ERPs.

Data and code availability statement

Data and code are available online (<https://osf.io/v59qu>).

Credit author statement

Zhenxing Hu: Conceptualization, Formal analysis, Methodology, Writing- Original draft preparation; **Zhiguo Zhang:** Conception, Study organization, Writing- Reviewing and Editing; **Zhen Liang:** Consultation, Data analysis; **Li Zhang:** Initial analysis; **Linling Li:** Data curation; **Gan Huang:** Supervision, Conceptualization, Methodology, Data collection, Writing- Reviewing and Editing

Acknowledgments

This work was supported by the National Natural Science Foundation of China (No. 81871443, 61906122, and 81901831), the Science, Technology, and Innovation Commission of Shenzhen Municipality Technology Fund (No. JCYJ20190808173819182), the Shenzhen Science and Technology Program (No. JSGG20210713091811038), the Shenzhen-Hong Kong Institute of Brain Science-Shenzhen Fundamental Research Institutions (No. 2021SHIBS0003).

None of the authors has potential conflicts of interest to be disclosed.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:[10.1016/j.neuroimage.2022.118937](https://doi.org/10.1016/j.neuroimage.2022.118937).

References

- Aron, A.R., Gluck, M.A., Poldrack, R.A., 2006. Long-term test-retest reliability of functional MRI in a classification learning task. Neuroimage 29, 1000–1006. doi:[10.1016/j.neuroimage.2005.08.010](https://doi.org/10.1016/j.neuroimage.2005.08.010).
- Boshra, R., Dhindsa, K., Boursalie, O., Ruiter, K.I., Sonnadara, R., Samavi, R., Doyle, T.E., Reilly, J.P., Connolly, J.F., 2019. From group-level statistics to single-subject prediction: machine learning detection of concussion in retired athletes. IEEE Trans. Neural Syst. Rehabil. Eng. 27, 1492–1501. doi:[10.1109/TNSRE.2019.2922553](https://doi.org/10.1109/TNSRE.2019.2922553).
- Brandmaier, A.M., Wenger, E., Bodammer, N.C., Kühn, S., Raz, N., Lindenberger, U., 2018. Assessing reliability in neuroimaging research through intra-class effect decomposition (ICED). Elife 7, e35718. doi:[10.7554/elife.35718.001](https://doi.org/10.7554/elife.35718.001).
- Bridgeford, E.W., Wang, S., Yang, Z., Wang, Z., Xu, T., Craddock, C., Dey, J., Kiar, G., Gray-Roncal, W., Coulantoni, C., 2020. Eliminating accidental deviations to minimize generalization error with applications in connectomics and genomics. Biorxiv 802629. doi:[10.1101/802629](https://doi.org/10.1101/802629).
- Cahn, B.R., Polich, J., 2006. Meditation states and traits: EEG, ERP, and neuroimaging studies. Psychol. Bull. 132, 180–211. doi:[10.1037/0033-2909.132.2.180](https://doi.org/10.1037/0033-2909.132.2.180).
- Cassidy, S.M., Robertson, I.H., O'Connell, R.G., 2012. Retest reliability of event-related potentials: evidence from a variety of paradigms. Psychophysiology 49, 659–664. doi:[10.1111/j.1469-8896.2011.01349.x](https://doi.org/10.1111/j.1469-8896.2011.01349.x).
- Croce, P., Quercia, A., Costa, S., Zappasodi, F., 2020. EEG microstates associated with intra- and inter-subject alpha variability. Sci. Rep. 10, 1–11. doi:[10.1038/s41598-020-58787-w](https://doi.org/10.1038/s41598-020-58787-w).
- Cruse, D., Beukema, S., Chennu, S., Malins, J.G., Owen, A.M., McRae, K., 2014. The reliability of the N400 in single subjects: implications for patients with disorders of consciousness. NeuroImage Clin 4, 788–799. doi:[10.1016/j.nicl.2014.05.001](https://doi.org/10.1016/j.nicl.2014.05.001).
- David, O., Harrison, L., Friston, K.J., 2005. Modelling event-related responses in the brain. Neuroimage 25, 756–770. doi:[10.1016/j.neuroimage.2004.12.030](https://doi.org/10.1016/j.neuroimage.2004.12.030).
- Devos, H., Burns, J.M., Liao, K., Ahmadnezhad, P., Mahnken, J.D., Brooks, W.M., Gustafson, K., 2020. Reliability of P3 event-related potential during working memory across the spectrum of cognitive aging. Front. Aging Neurosci. 12, 1–8. doi:[10.3389/fnagi.2020.566391](https://doi.org/10.3389/fnagi.2020.566391).
- Dubois, J., Adolphs, R., 2016. Building a science of individual differences from fMRI. Trends Cogn. Sci. 20, 425–443. doi:[10.1016/j.tics.2016.03.014](https://doi.org/10.1016/j.tics.2016.03.014).
- Fisher, A.J., Medaglia, J.D., Jeronimus, B.F., 2018. Lack of group-to-individual generalizability is a threat to human subjects research. Proc. Natl. Acad. Sci. U. S. A. 115, E6106–E6115. doi:[10.1073/pnas.1711978115](https://doi.org/10.1073/pnas.1711978115).
- Fliessbach, K., Rohe, T., Linder, N.S., Trautner, P., Elger, C.E., Weber, B., 2010. Retest reliability of reward-related BOLD signals. Neuroimage 50, 1168–1176. doi:[10.1016/j.neuroimage.2010.01.036](https://doi.org/10.1016/j.neuroimage.2010.01.036).
- Fröhner, J.H., Teckentrup, V., Smolka, M.N., Kroemer, N.B., 2019. Addressing the reliability fallacy in fMRI: similar group effects may arise from unreliable individual effects. NeuroImage 195, 174–189. doi:[10.1016/j.neuroimage.2019.03.053](https://doi.org/10.1016/j.neuroimage.2019.03.053).
- Frömer, R., Maier, M., Rahman, R.A., 2018. Group-level EEG-processing pipeline for flexible single trial-based analyses including linear mixed models. Front. Neurosci. 12, 1–15. doi:[10.3389/fnins.2018.00048](https://doi.org/10.3389/fnins.2018.00048).

- Gaspar, C.M., Rousselet, G.A., Pernet, C.R., 2011. Reliability of ERP and single-trial analyses. *Neuroimage* 58, 620–629. doi:[10.1016/j.neuroimage.2011.06.052](https://doi.org/10.1016/j.neuroimage.2011.06.052).
- Goodhew, S.C., Edwards, M., 2019. Translating experimental paradigms into individual-differences research: contributions, challenges, and practical recommendations. *Conscious. Cogn.* 69, 14–25. doi:[10.1016/j.concog.2019.01.008](https://doi.org/10.1016/j.concog.2019.01.008).
- Hedge, C., Powell, G., Sumner, P., 2018. The reliability paradox: why robust cognitive tasks do not produce reliable individual differences. *Behav. Res. Methods* 50, 1166–1186. doi:[10.3758/s13428-017-0935-1](https://doi.org/10.3758/s13428-017-0935-1).
- Höller, Y., Uhl, A., Bathke, A., Thomschewski, A., Butz, K., Nardone, R., Fell, J., Trinka, E., 2017. Reliability of EEG measures of interaction: a paradigm shift is needed to fight the reproducibility crisis. *Front. Hum. Neurosci.* 11, 1–15. doi:[10.3389/fnhum.2017.00441](https://doi.org/10.3389/fnhum.2017.00441).
- Hu, L., Iannetti, G.D., 2019. Neural indicators of perceptual variability of pain across species. *Proc. Natl. Acad. Sci. U. S. A.* 116, 1782–1791. doi:[10.1073/pnas.1812499116](https://doi.org/10.1073/pnas.1812499116).
- Huang, G., 2019. EEG/ERP data analysis toolboxes. In: Li, H., Zhiguo, Z. (Eds.), *EEG Signal Processing and Feature Extraction*. Springer, pp. 407–434. doi:[10.1007/978-981-13-9113-2](https://doi.org/10.1007/978-981-13-9113-2).
- Huffmeijer, R., Bakermans-Kranenburg, M.J., Alink, L.R.A., Van IJzendoorn, M.H., 2014. Reliability of event-related potentials: the influence of number of trials and electrodes. *Physiol. Behav.* 130, 13–22. doi:[10.1016/j.physbeh.2014.03.008](https://doi.org/10.1016/j.physbeh.2014.03.008).
- Infantolino, Z.P., Luking, K.R., Sauder, C.L., Curtin, J.J., Hajcak, G., 2018. Robust is not necessarily reliable: from within-subjects fMRI contrasts to between-subjects comparisons. *Neuroimage* 173, 146–152. doi:[10.1016/j.neuroimage.2018.02.024](https://doi.org/10.1016/j.neuroimage.2018.02.024).
- Ip, C.T., Ganz, M., Ozenne, B., Sluth, L.B., Gram, M., Viardot, G., l'Hostis, P., Danjou, P., Knudsen, G.M., Christensen, S.R., 2018. Pre-intervention test-retest reliability of EEG and ERP over four recording intervals. *Int. J. Psychophysiol.* 134, 30–43. doi:[10.1016/j.ijpsycho.2018.09.007](https://doi.org/10.1016/j.ijpsycho.2018.09.007).
- Jansen, B.H., Rit, V.G., 1995. Electrocorticogram and visual evoked potential generation in a mathematical model of coupled cortical columns. *Biol. Cybern.* 73, 357–366. doi:[10.1007/BF00199471](https://doi.org/10.1007/BF00199471).
- Kraemer, H.C., 2014. The reliability of clinical diagnoses: state of the art. *Annu. Rev. Clin. Psychol.* 10, 111–130. doi:[10.1146/annurev-clinpsy-032813-153739](https://doi.org/10.1146/annurev-clinpsy-032813-153739).
- Lee, K.H., Choi, Y.Y., Gray, J.R., Cho, S.H., Chae, J.H., Lee, S., Kim, K., 2006. Neural correlates of superior intelligence: stronger recruitment of posterior parietal cortex. *Neuroimage* 29, 578–586. doi:[10.1016/j.neuroimage.2005.07.036](https://doi.org/10.1016/j.neuroimage.2005.07.036).
- Leue, A., Klein, C., Lange, S., Beauducel, A., 2013. Inter-individual and intra-individual variability of the N2 component: on reliability and signal-to-noise ratio. *Brain Cogn.* 83, 61–71. doi:[10.1016/j.bandc.2013.06.009](https://doi.org/10.1016/j.bandc.2013.06.009).
- Li, M., Wang, D., Ren, J., Langs, G., Stoecklein, S., Brennan, B.P., Lu, J., Chen, H., Liu, H., 2019. Performing group-level functional image analyses based on homologous functional regions mapped in individuals. *PLoS Biol* 17, 1–27. doi:[10.1371/journal.pbio.2007032](https://doi.org/10.1371/journal.pbio.2007032).
- McGraw, K.O., Wong, S.P., 1996. Forming inferences about some intraclass correlations coefficients": correction. *Psychol. Methods* 1. doi:[10.1037//1082-989x.1.4.390](https://doi.org/10.1037//1082-989x.1.4.390), 390–390.
- Munsters, N.M., van Ravenswaaij, H., van den Boomen, C., Kemner, C., 2019. Test-retest reliability of infant event related potentials evoked by faces. *Neuropsychologia* 126, 20–26. doi:[10.1016/j.neuropsychologia.2017.03.030](https://doi.org/10.1016/j.neuropsychologia.2017.03.030).
- Nelson, E.E., Guyer, A.E., 2012. Beyond brain mapping: using neural measures to predict real-world outcomes. *Curr. Dir. Psychol. Sci.* 1, 233–245. doi:[10.1177/0963721412469394.Beyond](https://doi.org/10.1177/0963721412469394).
- Noble, S., Scheinost, D., Constable, R.T., 2019. A decade of test-retest reliability of functional connectivity: a systematic review and meta-analysis. *Neuroimage* 203, 116157. doi:[10.1016/j.neuroimage.2019.116157](https://doi.org/10.1016/j.neuroimage.2019.116157).
- Onie, S., Most, S., 2017. Two roads diverged: distinct mechanisms of attentional bias differentially predict negative affect and persistent negative thought. *Emotion* 17. doi:[10.1037/emo0000280](https://doi.org/10.1037/emo0000280).
- Pernet, C., Garrido, M.I., Gramfort, A., Maurits, N., Michel, C.M., Pang, E., Salmelin, R., Schoffelen, J.M., Valdes-Sosa, P.A., Puce, A., 2020. Issues and recommendations from the OHBM COBIDAS MEG committee for reproducible EEG and MEG research. *Nat. Neurosci.* 23, 1473–1483. doi:[10.1038/s41593-020-00709-0](https://doi.org/10.1038/s41593-020-00709-0).
- Plichta, M.M., Schwarz, A.J., Grimm, O., Morgen, K., Mier, D., Haddad, L., Gerdes, A.B.M., Sauer, C., Tost, H., Esslinger, C., Colman, P., Wilson, F., Kirsch, P., Meyer-Lindenberg, A., 2012. Test-retest reliability of evoked BOLD signals from a cognitive-emotive fMRI test battery. *Neuroimage* 60, 1746–1758. doi:[10.1016/j.neuroimage.2012.01.129](https://doi.org/10.1016/j.neuroimage.2012.01.129).
- Polich, J., 2004. Clinical application of the P300 event-related brain potential. *Phys. Med. Rehabil. Clin. N. Am.* 15, 133–161. doi:[10.1016/S1047-9651\(03\)00109-8](https://doi.org/10.1016/S1047-9651(03)00109-8).
- Raemaekers, M., Vink, M., Zandbelt, B., van Wezel, R.J.A., Kahn, R.S., Ramsey, N.F., 2007. Test-retest reliability of fMRI activation during prosaccades and antisaccades. *Neuroimage* 36, 532–542. doi:[10.1016/j.neuroimage.2007.03.061](https://doi.org/10.1016/j.neuroimage.2007.03.061).
- Segalowitz, S.J., Barnes, K.L., 1993. The reliability of ERP components in the auditory oddball paradigm. *Psychophysiology* 30, 451–459. doi:[10.1111/j.1469-8986.1993.tb02068.x](https://doi.org/10.1111/j.1469-8986.1993.tb02068.x).
- Seghier, M.L., Price, C.J., 2018. Interpreting and utilising intersubject variability in brain function. *Trends Cogn. Sci.* 22, 517–530. doi:[10.1016/j.tics.2018.03.003](https://doi.org/10.1016/j.tics.2018.03.003).
- Spearman, C., 1910. Correlation calculated from faulty data. *Br. J. Psychol.* 3, 271–295. doi:[10.1111/j.2044-8295.1910.tb00206.x](https://doi.org/10.1111/j.2044-8295.1910.tb00206.x), 1904–1920.
- Sur, S., Sinha, V.K., 2009. Event-related potential: an overview. *Ind. Psychiatry J.* 18, 70–73. doi:[10.4103/0972-6748.57865](https://doi.org/10.4103/0972-6748.57865).
- Thigpen, N.N., Kappenan, E.S., Keil, A., 2017. Assessing the internal consistency of the event-related potential: an example analysis. *Psychophysiology* 54, 123–138. doi:[10.1111/pisp.12629](https://doi.org/10.1111/pisp.12629).
- Tibon, R., Levy, D.A., 2015. Striking a balance: analyzing unbalanced event-related potential data. *Front. Psychol.* 6, 1–4. doi:[10.3389/fpsyg.2015.00555](https://doi.org/10.3389/fpsyg.2015.00555).
- Van Rijsbergen, N.J., Schyns, P.G., 2009. Dynamics of trimming the content of face representations for categorization in the brain. *PLoS Comput. Biol.* 5. doi:[10.1371/journal.pcbi.1000561](https://doi.org/10.1371/journal.pcbi.1000561).
- Vetter, N.C., Steding, J., Jurk, S., Ripke, S., Mennigen, E., Smolka, M.N., 2017. Reliability in adolescent fMRI within two years - a comparison of three tasks. *Sci. Rep.* 7, 1–11. doi:[10.1038/s41598-017-02334-7](https://doi.org/10.1038/s41598-017-02334-7).
- Yarkoni, T., Braver, T.S., 2010. Cognitive neuroscience approaches to individual differences in working memory and executive control: conceptual and methodological issues 87–107. https://doi.org/10.1007/978-1-4419-1210-7_6.
- Youssofzadeh, V., Prasad, G., Wong-Lin, K.F., 2015. On self-feedback connectivity in neural mass models applied to event-related potentials. *Neuroimage* 108, 364–376. doi:[10.1016/j.neuroimage.2014.12.067](https://doi.org/10.1016/j.neuroimage.2014.12.067).



Assessing reliability in neuroimaging research through intra-class effect decomposition (ICED)

Andreas M Brandmaier^{1,2*}, Elisabeth Wenger¹, Nils C Bodammer¹, Simone Kühn³, Naftali Raz^{1,4}, Ulman Lindenberger^{1,2}

¹Center for Lifespan Psychology, Max Planck Institute for Human Development, Berlin, Germany; ²Max Planck UCL Centre for Computational Psychiatry and Ageing Research, Berlin, Germany; ³Clinic and Policlinic for Psychiatry and Psychotherapy, University Clinic Hamburg-Eppendorf, Hamburg, Germany; ⁴Wayne State University, Detroit, United States

Abstract Magnetic resonance imaging has become an indispensable tool for studying associations of structural and functional properties of the brain with behavior in humans. However, generally recognized standards for assessing and reporting the reliability of these techniques are still lacking. Here, we introduce a new approach for assessing and reporting reliability, termed intra-class effect decomposition (ICED). ICED uses structural equation modeling of data from a repeated-measures design to decompose reliability into orthogonal sources of measurement error that are associated with different characteristics of the measurements, for example, session, day, or scanning site. This allows researchers to describe the magnitude of different error components, make inferences about error sources, and inform them in planning future studies. We apply ICED to published measurements of myelin content and resting state functional connectivity. These examples illustrate how longitudinal data can be leveraged separately or conjointly with cross-sectional data to obtain more precise estimates of reliability.

DOI: <https://doi.org/10.7554/eLife.35718.001>

*For correspondence:
brandmaier@mpib-berlin.mpg.de

Competing interests: The authors declare that no competing interests exist.

Funding: See page 17

Received: 06 February 2018

Accepted: 01 July 2018

Published: 02 July 2018

Reviewing editor: Heidi Johansen-Berg, University of Oxford, United Kingdom

© Copyright Brandmaier et al. This article is distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use and redistribution provided that the original author and source are credited.

Introduction

Neuroimaging techniques have become indispensable tools for studying associations among brain structure, brain function, and behavior in multiple contexts, including aging, child development, neuropathology and interventions, with concerted efforts increasingly focusing on comprehensive quantitative analyses across multiple imaging modalities ([Lerch et al., 2017](#)). Surprisingly, however, generally recognized standards and procedures for assessing and reporting the reliability of measurements and indices generated by noninvasive neuroimaging techniques are still lacking. This state of affairs may reflect the rapid evolution of a research field that straddles several well-established disciplines such as physics, biology, and psychology. Each of these fields comes with its own methodology, including conceptualization of error of measurement and reliability, and an articulation of these diverse methodologies into a coherent neuroscience framework is currently lacking. The goal of our contribution is two-fold. First, we introduce a signal-to-noise perspective that reconciles these seemingly disparate approaches. Second, we apply an analytic framework, based on the ideas of Generalizability Theory (G-Theory; [Cronbach et al., 1972](#)) and Structural Equation Modeling (SEM) that allows us to separate and gauge various sources of measurement error associated with different characteristics of the measurement, such as run, session, day, or scanning site (in multi-site studies). The proposed tool enables researchers to describe the magnitude of individual error components, make inferences about the error sources, and inform them in planning the design of future studies.

We proceed without loss of generality but with an emphasis on applications to human cognitive neuroscience.

Materials and methods

Prelude: Coefficient of variation and intra-class correlation coefficient represent different but compatible conceptions of signal and noise

Physics and psychometrics offer two fundamentally different but equally important and compatible conceptions of reliability and error. Physicists typically inquire how reliably a given measurement instrument can detect a given quantity. To this end, they repeatedly measure a property of an object, be it a phantom or a single research participant, and for expressing the absolute precision of measurement, evaluate the dispersion of the different measurement values obtained from this object to their mean. The prototypical index produced by such approach is the coefficient of variation (CV), which is defined as the ratio of an estimate of variability, σ_i , and a mean, m , with i representing the object undergoing repeated measurements:

$$CV_i = \frac{\sigma_i}{m_i}$$

The interpretability of the CV depends upon the quantity having positive values and being measured on a ratio scale. When these conditions are met, the CV effectively expresses the (im)precision of measurement, with larger values meaning lesser precision. Imagine, for instance, that the same quantity is being measured in the same research participant or the same phantom on two different scanners. All other things equal, comparing the CV obtained from each of the two scanners shows which of the two provides a more reliable (in this case, precise) measurement.

Note that in this context, the scanner with the greater precision may not necessarily yield more valid data, as the mean of its measurements may be further away from the ground truth (see *Figure 1*). Bearing this distinction in mind, we limit our discussion to the issues of reliability (precision), rather than validity (bias). In *Table 1*, we list terms used in various disciplines to express the difference between precision and bias. We maintain that the confusion surrounding these concepts may to a large extent reflect terminological differences among disciplines.

In contrast to physics that deals with well-defined objects of measurement, in human neuroscience, we focus on a different meaning of reliability. Informed by psychometric theory and differential psychology, reliability here refers to the precision of assessing between-person differences. Researchers concerned with gauging individual differences as a meaningful objective express this form of reliability in a ratio index, termed intra-class correlation coefficient (ICC), which represents the strength of association between any pair of measurements made on the same object. However, instead of relating variance to the mean, the ICC quantifies variance *within* persons (or groups of persons), in relation to the total variance, which also contains variance *between* persons (or between groups of persons; cf. *Bartko, 1966*). Hence, the ICC is a dimensionless quantity bracketed between 0 and 1, and is tantamount to the ratio of variance-between, σ_B^2 , to the total variance that includes the variance-within, σ_W^2 :

$$ICC = \frac{\sigma_B^2}{\sigma_W^2 + \sigma_B^2}$$

In repeated-measures studies on human participants, the variance-within corresponds to the variance within each person, whereas the variance-between represents differences among persons. Thus, for interval or ratio scales, the ICC expresses the percentage of the total variance that can be attributed to differences between persons.

The similarities and differences between CV and ICC become clear when one conceives of both as expressions of signal-to-noise ratio. For a physicist, the mean represents the sought-after signal, and the variation around the mean represents the noise to be minimized. Hence the use of the CV to evaluate measurement precision normalized with the metric of the given scale. For a psychologist interested in individual differences, the between-person variation is the signal, and the within-person variation is regarded as noise. Therefore, a measure that quantifies the contribution of between-person differences to the total variance in the data, the ICC, is chosen for this purpose (in other

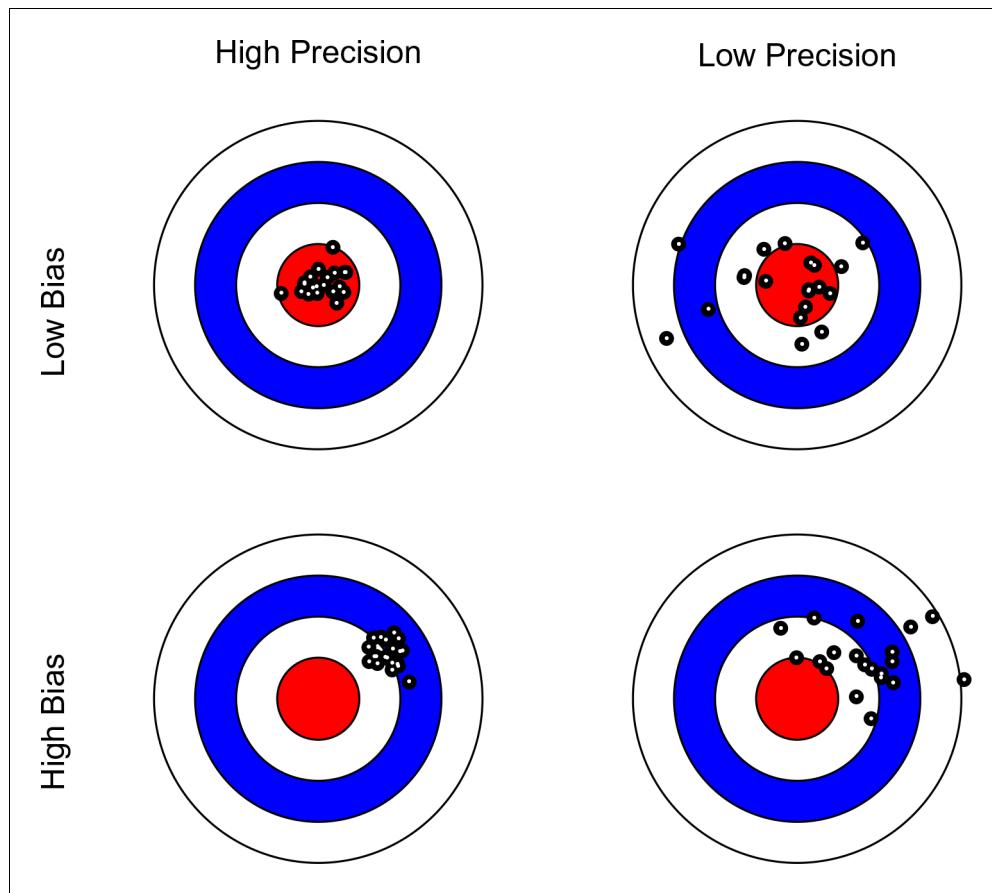


Figure 1. Bullseye charts representing precision and bias of a measurement instrument. The center of each bullseye represents ground truth and the black dots represent repeated measurements. Ideal measurement instruments have minimal bias and maximal precision (as illustrated in the top-left panel).

DOI: <https://doi.org/10.7554/eLife.35718.002>

contexts, not discussed in this article, within-person variability itself may be an important marker of individual differences, e.g., *Garrett et al., 2013; Nesselroade, 1991*.

Clearly, CV and ICC do not convey the same information. To illustrate this point, we simulated data under two conditions, which show that each measure can be manipulated independently of the other. We illustrate how CV remains unchanged, while drastic changes occur in ICC (see **Figure 2**). Instead of individual CV values, we report an aggregated CV computed as the square-root of the average within-person variance divided by the overall mean. For each condition, we simulated for each of five persons ten repeated measures of a fictitious continuous outcome variable X . Across conditions and persons, within-person variability was identical and only between-person variability

Table 1. A list of terms describing the concepts of variability across repetitions and average distance from ground truth over measurements across different disciplines and knowledge domains.

Knowledge domain	Variability across repetitions	Average distance from ground truth
Psychology	Reliability	Validity
Physics	Precision	Accuracy
Statistics	Variance	Bias
Measurement Theory	Random error	Systematic error
ISO 5725 ('Accuracy')	Precision	Trueness

DOI: <https://doi.org/10.7554/eLife.35718.003>

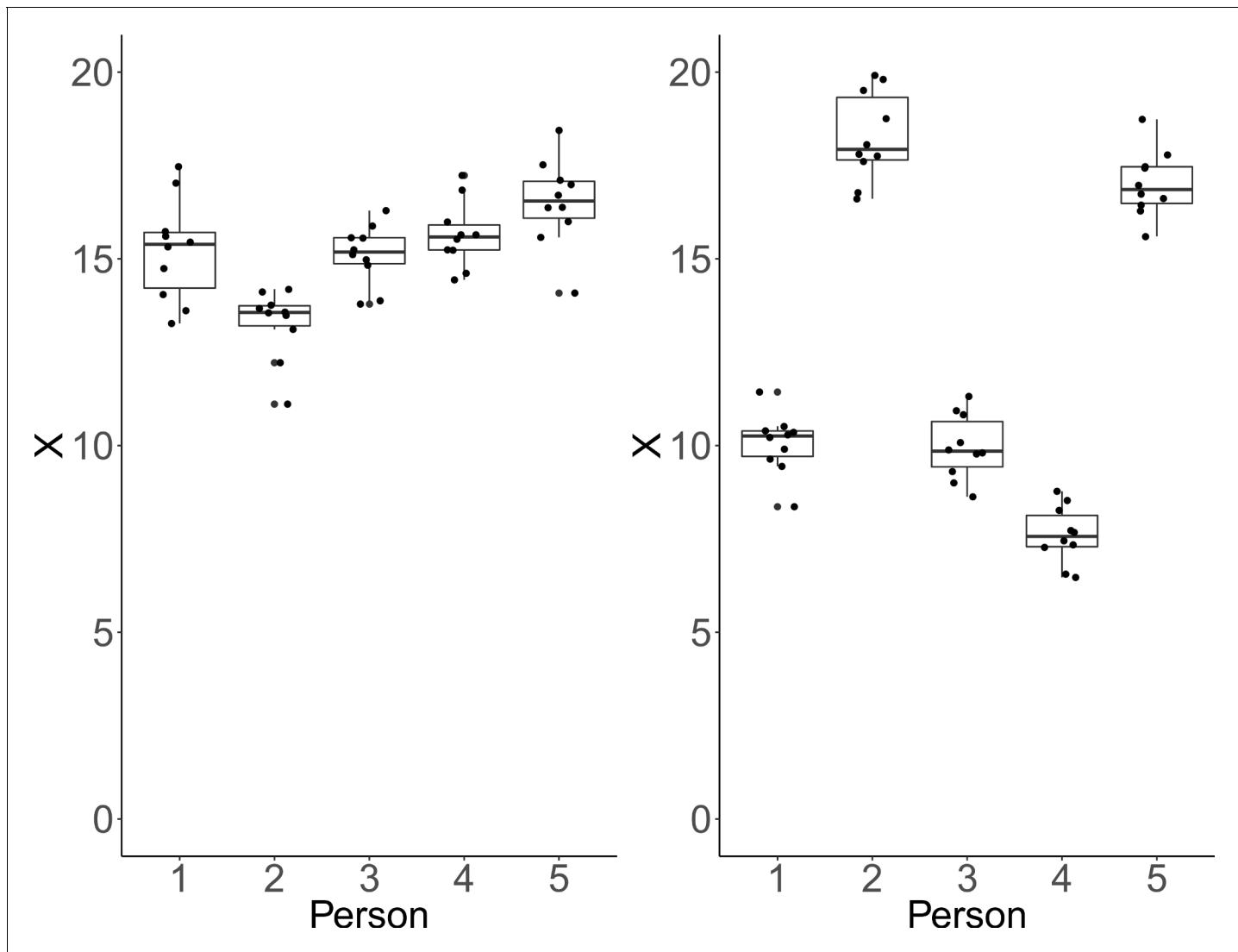


Figure 2. Each plot shows simulated data of 10 repeated measurements of a fictitious outcome X for five persons. Within-person variance is identical over persons and panels. Left panel: Between-person variance is identical to within-person variance. In that case, ICC-indexed reliability is estimated at a rather low $ICC = 0.48$, whereas CV-indexed precision is fair, with averaged $CV = 0.07$. Right panel: Between-person standard deviation is larger than within-person standard deviation by a factor of five. In this condition, estimated ICC-indexed reliability is high ($ICC = 0.96$), but CV-indexed precision remains identical with averaged $CV = 0.07$ as the within-person standard deviation is still fairly small with respect to the sample mean.

DOI: <https://doi.org/10.7554/eLife.35718.004>

varied between conditions. In the first condition, the simulated data have identical between-person and within-person variance. As a result, we obtain a low ICC and conclude that the measurement instrument fails to adequately discriminate among persons. However, critically, we also obtain a rather low CV , implying high precision to detect deviation from zero (see left panel of **Figure 2**). In the second instance, between-person standard deviation was larger than within-person standard deviation by a factor of five. This condition yields a high ICC reflecting the fact that the measure discriminates well among persons. At the same time, CV remains low, which implies reasonable precision of detecting differences from zero. This is because the within-person variance is still relatively low in comparison to the means (see right panel of **Figure 2**).

In summary, whereas the CV refers to the precision of measurement obtained from each object, the ICC expresses a part-whole relation of variance observed in the data. All other things being equal, a less precise measurement will increase the variance-within, and hence compromise our ability to detect between-person differences. On the other hand, a rather imprecise measurement (as

indexed by the CV obtained for each object of measurement) may nevertheless yield high reliability (as indexed by the ICC) if the between-person differences in means are large.

Intra-class effect decomposition (ICED)

The extant neuroimaging literature typically offers little justification for the choice of the reliability index. Based on the preceding considerations, this is problematic, as the various indices differ greatly in meaning. The ICC and variants thereof are appropriate for evaluating how well one can capture between-person differences in a measure of interest. Put differently, it is misleading to report the CV as a measure of reliability when the goal of the research is to investigate individual or group differences. Both approaches to reliability assessment are informative, but they serve different purposes, and cannot be used interchangeably. Below, we focus on individual differences as we present a general and versatile method for estimating the relative contributions of different sources of error in repeated-measures designs. Our approach can be seen as an extension of ANOVA-based approaches to decomposing ICCs. In this sense, it is tightly linked to G-Theory, which has been used successfully before in assessing reliability of neuroimaging measures (Gee et al., 2015; Noble et al., 2017). The method, termed intra-class effect decomposition (ICED), has ICC as its core concept. The key feature of the method, however, is its ability to distinguish among *multiple* sources of unreliability, with the understanding that not all sources of error and their separation are important and meaningful in repeated-measures designs. For example, different sources of error may be due to run, session, day, site, scanner, or acquisition protocol variations. Furthermore, there may be more complex error structures to be accounted for, for example, runs nested in sessions; and multiple sessions, again, may be nested within days, and all may be nested under specific scanners in multi-site investigations. Neglecting these nuances of error structures leads to biased reliability estimates. The ability to adequately model these relationships and visually represent them in path diagrams is a virtue of our approach.

Beyond reliability per se, researchers may often be interested in the specific sources of error variance and measurement characteristics that contribute to it. For example, in applying MRI to studying long-term within-person changes in the course of aging, child development, disease progression, or treatment, one may wish to determine first what effect repositioning of a person in the scanner between sessions has on reliability of measured quantities (e.g., Arshad et al., 2017). Similarly, it may be important to determine how much variation is associated with scanning on a different day relative to conducting two scanning sessions on the same day (e.g., Morey et al., 2010). These types of questions are of utmost importance in longitudinal studies, in which researchers collect data on the same person using an ostensibly identical instrument (e.g., MRI scanner) under an identical protocol (sequence), but inevitably under slightly different measurement characteristics, including position of the participant within the scanner, body and air temperature, or time of day. From a design perspective, knowing the distinct components of measurement error and their relative magnitudes may enhance future study designs and boost their generalizability.

In the proposed SEM framework, observed variance is partitioned into several orthogonal error variance components that capture unreliability attributable to specific measurement characteristics, with the number of components depending on identification constraints based on the study design. **Figure 3** shows a minimal, or optimally efficient, repeated-measures study design for estimating the contributions of the main effects of day, session, and residual variance to measurement error. The design consists of four measurements (scans) performed over two days and three sessions. In this design, unique contributions of each error source are identified as depicted in the path diagram in **Figure 4**. In the diagram, observed variables correspond to image acquisitions and are depicted as rectangles; latent variables are depicted as circles and represent the unobservable sources of variance, that is, the true score variance (T) and the error variance components of day (D), session (S), and residual (E). Double-headed arrows represent variances of a latent variable. Single-headed arrows represent regressions with fixed unit loadings.

In this example, total observed variability in an outcome across measurements and persons is partitioned into true-score variance and three error variance terms: the day-specific error variance, the session-specific error variance (here capturing the effect of repositioning a person between scans), and the residual error variance. The full measurement model is depicted as a path diagram in the left panel of **Figure 4**. The structural equation model specifies four observed variables representing the repeated measurements of the outcome of interest. One of the latent variables represents the

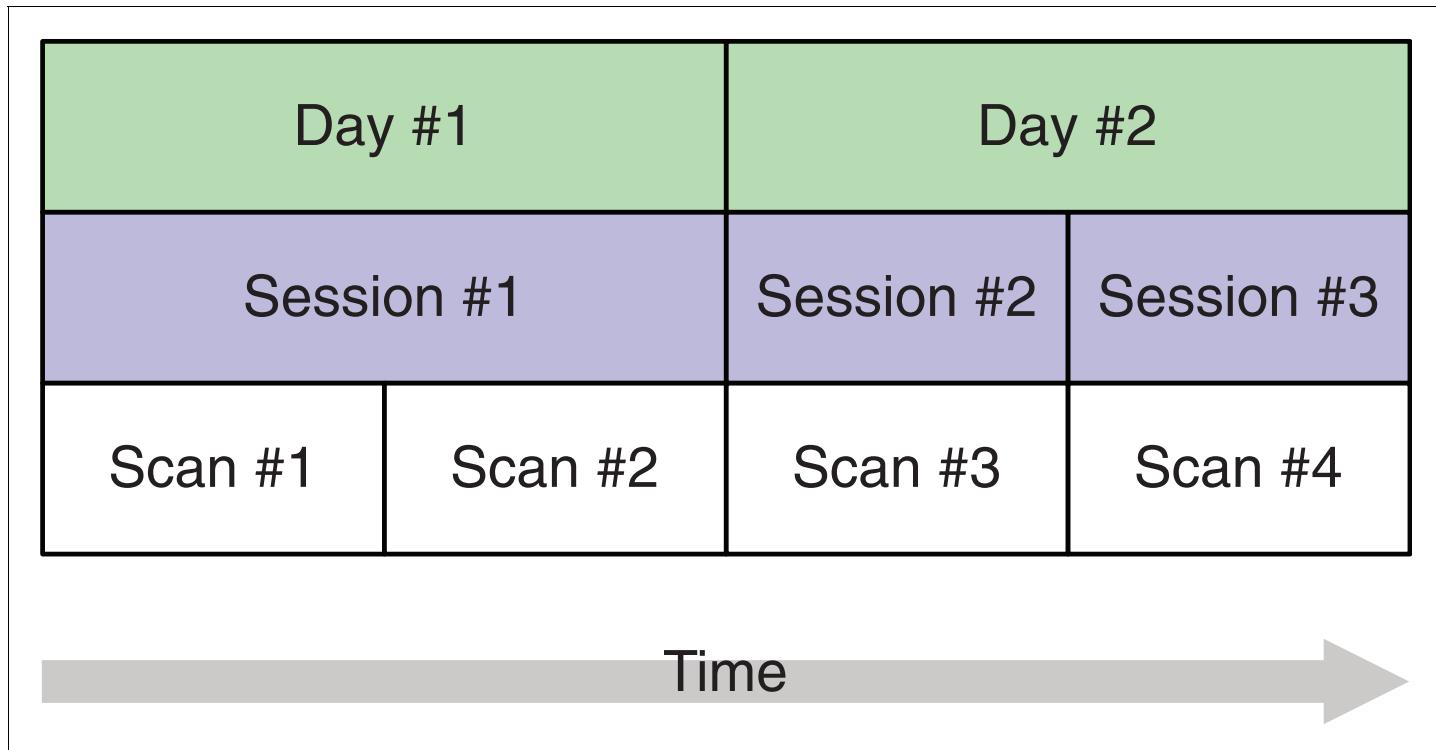


Figure 3. A study design with four brain imaging scans per person spread across three sessions on two days. The start of each session (blue) implies that the person is moved into the scanner. On the first day, there is only a single session, that is, between scans 1 and 2, the person remains in the scanner whereas on day 2, the person is removed from the scanner after the first session and, after a short break, placed back in. This allows separating the session-specific and the day-specific variance contributions to total variance.

DOI: <https://doi.org/10.7554/eLife.35718.005>

true values of the construct of interest. Its variance, σ_T^2 , denotes the between-person variance. Fixed regressions of each measurement occasion on the latent construct express the assumption that we are measuring the given construct with each of the four repeated measures on the same scale. There are four orthogonal error variance sources with identical residual variance σ_E^2 , that is, residual errors that are not correlated with any other type of error or among themselves over time. In classical test theory, this is referred to as a *parallel model*, in which the construct is measured on the same scale with identical precision at each occasion. Typically, there is no explicit assumption of uncorrelated error terms even though many measures derived from this theory assume (and are only valid under) uncorrelated error terms (Raykov et al., 2015). Here, we focus on a parallel model while accounting for the correlated error structure implied by the greater similarity of multiple runs within the same session compared to runs across different sessions. Note that, in the SEM framework, we also can extend the parallel model to more complex types of measurement models (e.g., con-generic or tau-equivalent models) that allow for different residual error variances or different factor loadings. To account for the nested structure in our design, we introduce two day-specific error variance sources with variance σ_D^2 that represent day-specific disturbances and imply a closer similarity of measurements on the same day. Finally, there are three session-specific variance sources (depicted in blue) representing the session effect (including, for example, the effect of repositioning a person between sessions). The model-implied covariance matrix has the total variances for each observed variable in the diagonal. It can be analytically or numerically derived using matrix algebra (McArdle, 1980) or path-tracing rules (Boker et al., 2002), and is typically available in SEM computer programs (e.g., von Oertzen et al., 2015). The full model-implied covariance matrix is given in Table 2. For the given study design, each variance source is uniquely identifiable, as there is a unique solution for all parameters in the model. From the covariance matrix, it is apparent that the inclusion of the session variance term differentially affects the similarity of measurements between days 1 and 2. The

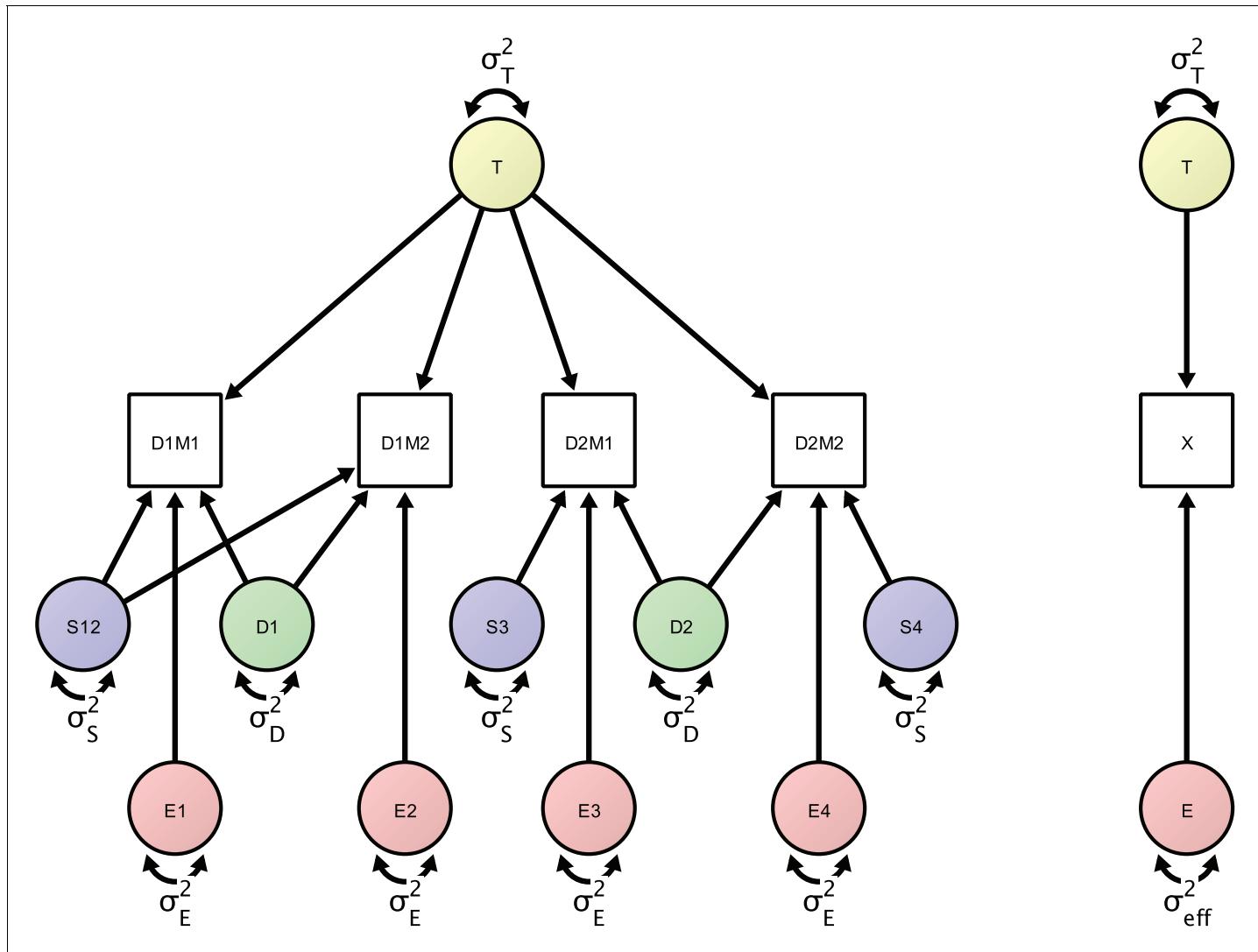


Figure 4. Left: Measurement model of four repeated measures according to the study design protocol. The measurement model includes day-specific effects that are separated by the error variance sources (in green), session-specific effects (represented by error variance sources in blue) and orthogonal residual error variance sources (represented in red). Right: A minimal, power-equivalent model that represents a direct measurement of the construct of interest. The true score (representing the outcome of interest) is measured with only a single error construct, whose variance is the effective error variance representing the combined influence of the complete error (co)variance structure shown in the left model.

DOI: <https://doi.org/10.7554/eLife.35718.006>

correlation between first and second scan is $\frac{\sigma_T^2 + \sigma_D^2 + \sigma_S^2}{\sigma_T^2 + \sigma_E^2 + \sigma_D^2 + \sigma_S^2}$ whereas the correlation between measurements 3 and 4 is $\frac{\sigma_T^2 + \sigma_D^2}{\sigma_T^2 + \sigma_E^2 + \sigma_D^2 + \sigma_S^2}$. Thus the similarity of the two measurements on the first day is greater than the similarity of measurements on the second day. In other words, the difference in correlation is the proportion of variance that the session-specific variance accounts for in total variance.

For this model (see **Figure 4**), we define ICC equivalently to the common ICC formula as ratio of between-person variance to total variance at the level of observed variables:

$$ICC = \frac{\sigma_T^2}{\sigma_T^2 + \sigma_E^2 + \sigma_D^2 + \sigma_S^2}$$

We estimate the components using the full information maximum likelihood procedure for SEM (**Finkbeiner, 1979**), which allows estimating all components under the assumption of the data

Table 2 Model-implied covariance matrix.

Rows and columns correspond to the four measurement occasions (brain scans) distributed over two days. Parameters of the covariance matrix are the individual differences of the construct of interest, σ_T^2 , the session-specific error variance, σ_S^2 , the day-specific error variance, σ_D^2 , and the residual error variance, σ_E^2 .

	Scan 1	Scan 2	Scan 3	Scan 4
Scan 1	$\sigma_T^2 + \sigma_E^2 + \sigma_D^2 + \sigma_S^2$	$\sigma_T^2 + \sigma_D^2 + \sigma_S^2$	σ_T^2	σ_T^2
Scan 2	$\sigma_T^2 + \sigma_D^2 + \sigma_S^2$	$\sigma_T^2 + \sigma_E^2 + \sigma_D^2 + \sigma_S^2$	σ_T^2	σ_T^2
Scan 3	σ_T^2	σ_T^2	$\sigma_T^2 + \sigma_E^2 + \sigma_D^2 + \sigma_S^2$	$\sigma_T^2 + \sigma_D^2$
Scan 4	σ_T^2	σ_T^2	$\sigma_T^2 + \sigma_D^2$	$\sigma_T^2 + \sigma_E^2 + \sigma_D^2 + \sigma_S^2$

DOI: <https://doi.org/10.7554/eLife.35718.007>

missing at random. This maximum-likelihood-based ICC is similar to the analytical procedure based on relating the ANOVA-derived within and between residual-sums-of-squares. The main difference is that the maximum likelihood estimator cannot attain negative values when we allow only positive variance estimates (Pannunzi et al., 2018).

In many cognitive neuroscience studies, one may be interested in construct-level reliability, and not only in reliability of indicators (i.e., observed variables). This construct reliability is captured by ICC_2 (Bliese, 2000). Based on the above SEM-based effect decomposition, we use power equivalence theory (von Oertzen, 2010) to derive the effective error of measuring the latent construct of interest. The effective error can be regarded as the residual error that would emerge from a direct measurement of a latent construct of interest. Here, it is an index of the precision with which a given study design is able to capture stable individual differences in the outcome of interest. The effective error is a function of all error components and its specific composition depends on the specific design in question. Effective error is the single residual error term that arises from all variances components other than the construct that is to be measured. As such, it represents the combined influence of all error variance components that determine construct reliability:

$$ICC_2 = \frac{\sigma_T^2}{\sigma_T^2 + \sigma_{eff}^2}$$

Effective error can be computed using the algorithm provided by von Oertzen (2010) and for some models, analytic expressions are available (see the multi-indicator theorem in von Oertzen, 2010). For the study design in our example, effective error is:

$$\sigma_{eff}^2 = \frac{(2\sigma_D^2 + \sigma_E^2 + \sigma_S^2)(2\sigma_D^2 + \sigma_E^2 + 2\sigma_S^2)}{8\sigma_D^2 + 4\sigma_E^2 + 3\sigma_S^2}$$

Relating true score variance to total variance yields ICC_2 – a measure of reliability on the construct level. For our model, ICC_2 is then:

$$ICC_2 = \frac{\sigma_T^2}{\sigma_T^2 + \frac{(2\sigma_D^2 + \sigma_E^2 + \sigma_S^2)(2\sigma_D^2 + \sigma_E^2 + 2\sigma_S^2)}{(8\sigma_D^2 + 4\sigma_E^2 + 3\sigma_S^2)}}$$

As a check, when assuming no day-specific and session-specific effects by inserting $\sigma_D^2 = 0$ and $\sigma_S^2 = 0$, we obtain the classical definition of ICC_2 that scales residual error variance with the number of measurement occasions (here, four occasions):

$$ICC_2 = \frac{\sigma_T^2}{\sigma_T^2 + \frac{\sigma_E^2}{4}}$$

In sum, ICC is a coefficient describing test-retest reliability of a measure (also referred to as short-term reliability or intra-session reliability by Noble et al., 2017) whereas ICC_2 is a coefficient

describing test-retest reliability of an underlying construct (an average score in parallel models) in a repeated-measures design (long-term reliability or intersession reliability according to **Noble et al., 2017**).

For our hypothesized measurement model that includes multiple measurements and multiple variance sources, the analytic solution of ICC_2 allows, for instance, to analytically trace reliability curves depending on properties of a design, such as the number of sessions, number of runs per sessions, number of sessions per day, or varying magnitudes of the error component. Of note, this corresponds to a D-study in G-Theory that can demonstrate, for example, how total session duration and number of sessions influence resting state functional connectivity reliability (see **Noble et al., 2017**).

A virtue of the proposed SEM approach is the possibility of applying likelihood-ratio tests to efficiently test simple and complex hypotheses about the design. For example, we can assess whether individual variance components significantly differ from zero or from particular values, or whether variance components have identical contributions (corresponding to F-tests on variance components in classical G-Theory). Such likelihood-ratio tests represent statistical model comparisons between a full model, in which each of the hypothesized error components are freely estimated from the data, and a restricted model, in which the variance of a target error component is set to zero. Both models are nested, and under the null hypothesis, the difference in negative-two log-likelihoods of the models will be χ^2 -distributed with 1 degree of freedom. This allows the derivation of *p* values for the null hypotheses of each individual error component being zero. Moreover, the generality of SEM allows testing complex hypotheses with hierarchically nested error structures or multi-group models. It also allows inference under missing data or by evaluating informative hypotheses (**de Schoot et al., 2011**) whereas ANOVA-based approaches become progressively invalid with increasing design complexity.

Results

An empirical example: Myelin water fraction data from Arshad et al. (2017)

To demonstrate how the proposed approach separates and quantifies sources of un-reliability, we re-analyzed data from a study of the brain regional myelin content by **Arshad et al. (2017)**. In human aging, changes of myelin structure and quantity have been proposed as neuroanatomical substrates of cognitive decline, which makes it particularly interesting to obtain a highly reliable estimate of regional myelin content, here, represented by myelin water fraction (MWF) derived from multi-component T_2 relaxation curves. The data in this demonstration were collected in 20 healthy adults (mean age \pm SD = 45.9 ± 17.1 years, range of 24.4–69.5 years; no significant difference between men and women: $t(18)=-0.81$, $p=0.43$) and are freely available (**Arshad et al., 2018**); for detailed sample description see **Arshad et al. (2017)**. The study protocol stipulated three acquisitions for each participant in a single session. In the first part, T_1 -weighted and T_2 -weighted MRI images were acquired, followed by a back-to-back acquisition of the ME- T_2 relaxation images without repositioning the participant in the scanner. At the end of the first part, participants were removed from the scanner and, after a short break, placed back in. In the second part, T_1 -weighted, T_2 -weighted and ME- T_2 multiecho sequences were acquired once. All further details relating to the study design, MR acquisition protocol, and preprocessing can be found in the original publication by **Arshad et al. (2017)**. In the following, we focus on the MWF derived from a multi-echo gradient recall and spin-echo (GRASE) sequence. The study design allows separating the influences of repositioning expressed as session-specific variance from true score variance (defined as the shared variance over all three repetitions) and individual error variance (the orthogonal residual error structure). **Figure 5** presents a diagram of the hypothesized contributions of the individual variance components. Parameters in the SEM correspond to estimates of true score variation (T), session-specific error variance component (S), and a residual error variance component (E). Model specification and estimation was both performed in Ω nyx (**von Oertzen et al., 2015**) and lavaan (**Rosseel, 2012**) via full information maximum likelihood. We provide the Ω nyx models and lavaan syntax in the Supplementary material.

For illustration, we only report estimates of the first of the six regions of interest reported in the original study, the anterior limb of the internal capsule (ALIC). The estimates of the individual variance components explaining the observed variance are shown in the diagram in **Figure 5**. To assess

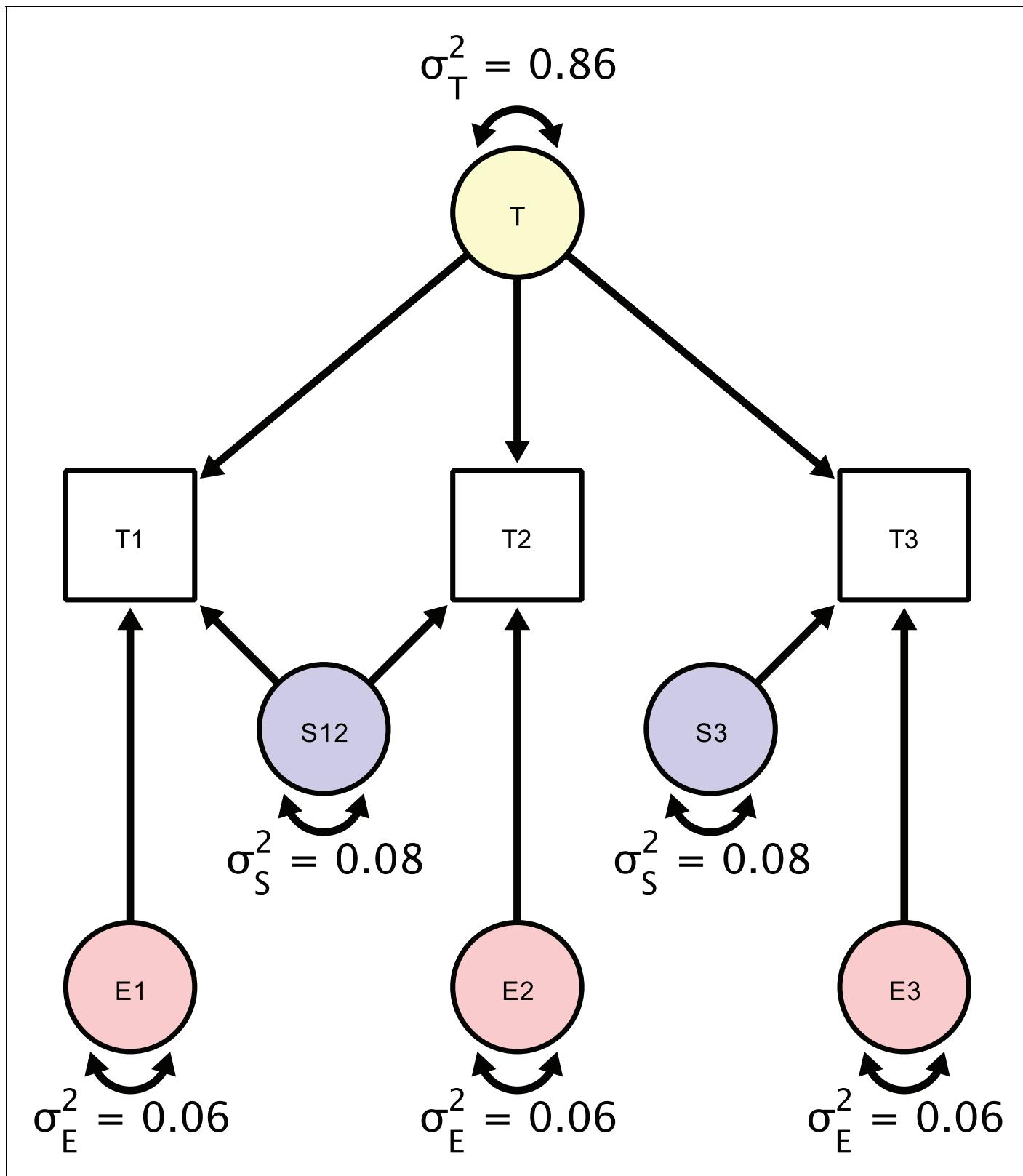


Figure 5. A Structural Equation Model of a repeated measures design in which each participant is scanned three times. Each person is scanned the first time (T1), followed by a back-to-back immediate re-acquisition (T2), and, finally, the person is moved out of the scanner, positioned back in the scanner, and scanned the third time (T3). Parameters in the SEM correspond to estimates of true score variation (σ_T^2), session-specific error variance

Figure 5 continued on next page

Figure 5 continued

component (σ_S^2), and a residual error variance component (σ_E^2). The standardized estimates are based on myelin water fraction data in the anterior limb of the internal capsule acquired from **Arshad et al. (2017)**.

DOI: <https://doi.org/10.7554/eLife.35718.008>

the significance of these components' magnitudes, we used likelihood ratio tests against null models, in which each component's variance was set to zero. For testing the residual error variance component, we used a Wald test because the null model without an orthogonal error structure cannot be estimated. Of the total between-person variance in measurements of MWF, we found that 86% were due to true score variance (est = 6.97; $\chi^2 = 27.759$; df = 1; p < 0.001), 8% - to session-specific variance (est = 0.59, $\chi^2 = 3.951$; df = 1; p = 0.047), and 6% - to residual error variance (est = 0.52; Z = 9.64; p = 0.002). Testing whether the variance contribution of the session-specific variance and the residual error variance were equal yields a non-significant result ($\chi^2 = 20.02$; df = 1; p = 0.89) and, thus, cannot be decided.

As shown before, we can obtain ICC as the ratio of systematic (t) and all variance components, which by means of standardization of the observed variables sums up to unity, resulting in:

$$ICC = 0.86$$

To compute ICC₂ as a standardized estimate of the precision with which the repeated-measures study design can measure individual differences in MWF in ALIC, we equate the day-specific variance with zero since it is not identified in this design but rather subsumed under the estimate of the true-score variance component, yielding:

$$ICC_2 = \frac{\sigma_T^2}{\sigma_T^2 + \frac{(\sigma_E^2 + \sigma_S^2)(\sigma_E^2 + 2\sigma_S^2)}{(3\sigma_E^2 + 4\sigma_S^2)}} = 0.94$$

The fact that day-specific variance and true score variance are inseparable in this design (both are shared variance components of all three measurement occasions) leads to an inflation of the true score variance estimate if non-zero day-specific variance is assumed and, thus, to an overly optimistic estimate of reliability. To be able to separate the individual variance contributions, one would have to rely on an augmented design that includes additional scanner acquisitions on at least one different day, such as the design shown in **Figure 4**.

Arshad et al. (2017) only reported pair-wise ICCs, based either only on the two back-to-back sessions of a single day, or on a single session of each day (again omitting a third of the available data). In the following, we derive the corresponding pair-wise ICCs using the full data set. Our estimates are similar even though not identical to the results obtained by **Arshad et al. (2017)** because our results were jointly estimated from three measurements. First, the authors report an estimate of ICC based on one measurement from the second session of the first day and one measurement from the single session of the second day, resulting in ICC = 0.83, which is close to our estimate of ICC = 0.86. Second, they reported an estimated reliability (ICC) derived only from the two back-to-back sessions on the first day as ICC = 0.94. Similarly, we can derive the reliability of a single measurement, had we measured only the two back-to-back sessions, achieving the identical result:

$$ICC = \frac{\sigma_T^2 + \sigma_S^2}{\sigma_T^2 + \sigma_S^2 + \sigma_E^2} = 0.94$$

The estimates of construct-level reliability obtained imply that individual differences in MWF can be measured quite well. As expected, the reliability estimate is higher for the back-to-back session than for the complete design because one error variance component, session-specific error variance, is not apportioned to the total error variance. Such a simple design commingles true score variance and the session-specific variance, and reliability studies should thus, by design, take into account potential different error sources, such as session-specific error variance.

A comprehensive SEM approach to assessing reliability allows for using the complete dataset in a single model to estimate reliability as either item-level reliability (ICC) or a construct-level reliability (ICC₂). A particular benefit of the proposed approach is its ability to tease apart individual error

components as far as the study design permits this, that is, as far as these components are identified. Future studies may very well increase study design complexity to test for additional error variance components. To compare the effect of repositioning a participant versus scanning a participant back-to-back, **Arshad et al. (2017)** compared pairwise ICCs of either the two back-to-back acquisitions or the second of the back-to-back acquisition with the repositioned acquisition. Using the SEM-based approach described above, we can directly estimate a variance component that quantifies the contribution of the session to the total error variance. We can also formally test whether this contribution is non-zero, or, if necessary, whether it is greater than some value or some other error variance component in the model. Furthermore, our estimates are always based on the complete dataset and there is no need to select certain pairs of runs for computing subset ICCs and potentially disregarding important dependencies in the data – a limitation that **Arshad et al. (2017)** explicitly mentioned in their report.

Link-wise reliability of resting state functional-connectivity indices

Resting-state functional connectivity was proposed as a promising index of age-related or pathology-induced changes in the brain, and has been used to predict brain maturation (**Dosenbach et al., 2010**) or disease state (**Craddock et al., 2009**). These applications can only prove practically useful if reliability is sufficiently high, so that differences between persons can be reliably detected in the first place, as a methodological precondition for prediction. Thus, there has been increasing interest in examining reliability of methods for assessing resting state connectivity (**Gordon et al., 2017; Noble et al., 2017; Pannunzi et al., 2018**). Here, we demonstrate how ICED can be used to evaluate reliability of pairwise functional indices obtained from resting-state functional connectivity analyses.

To illustrate such a model, we obtained the resting state functional connectivity (rsFC) dataset from **Pannunzi et al., 2018**, which is based on the publicly available raw data from the Day2day study (**Filevich et al., 2017**). In that study, six participants were scanned at least 40 (and some up to 50) times over the course of approximately seven months, and another sample of 50 participants (data from 42 participants of them available) were each scanned only once. In the following, we show how both datasets can be jointly investigated to estimate link-wise reliability of resting state functional connectivity (rsFC). We present a reliability analysis of the link-wise connectivity indices of brain regions-of-interest based on 5 min of measurement. For each measurement, as our main outcome, we obtained a 16×16 correlation matrix of rsFC indices, for pairs of regions including pre-frontal, sensor-motor, parietal, temporal, limbic, occipital cortices, cerebellum and subcortical structures. In our model, we assume independence of the measurement occasions. Thus, we decompose the covariance structure of the repeated measurements into one between-person variance and one within-person variance component. For simplicity, we illustrate this model by using the first ten observations. **Figure 6** shows a path diagram of this model. We estimated this model using **Omega** and **lavaan** and significance tests were performed using Wald tests. For example, we first estimated our model only for the link between left prefrontal cortex and right prefrontal cortex. The true score variance was estimated to account for 49% of the total variance (est = 0.013; W = 2.46; df = 1; p=0.117) and the error variance contributed 51% of the total variance (est = 0.014; W = 27.00; df = 1; p<0.0001), thus, ICC was 0.49.

With up to fifty measurement occasions, we can expect to get sufficiently precise measures of within-person fluctuations but since only eight participants contributed, we augment this dataset with cross-sectional data from additional 42 persons treating them as quasi-longitudinal data with the majority of data missing. This more precise measurement of between-person differences yields a somewhat different pattern of results. The true score variance was 39% of the total variance (est = 0.008; W = 6.31; df = 1; p=0.012) and the error variance was 61% of the total variance (est = 0.013; W = 33.53; df = 1; p<0.0001). Thus, our estimate dropped from 0.49 to 0.39. Due to a small sample size in the first analysis, we likely had overestimated the between-person differences in rsFC and had obtained an exceedingly overoptimistic ICC. By augmenting the initial analysis with a second dataset, we have obtained more precise and, here, even more pessimistic estimates of rsFC reliability.

Figure 7 shows a reliability matrix of all links between the investigated brain regions with estimates based on the joint model. **Pannunzi et al., 2018** reported that ICCs range from 0.0 to 0.7 with an average ICC of 0.22, which is typically considered an unacceptably low reliability (i.e., signal

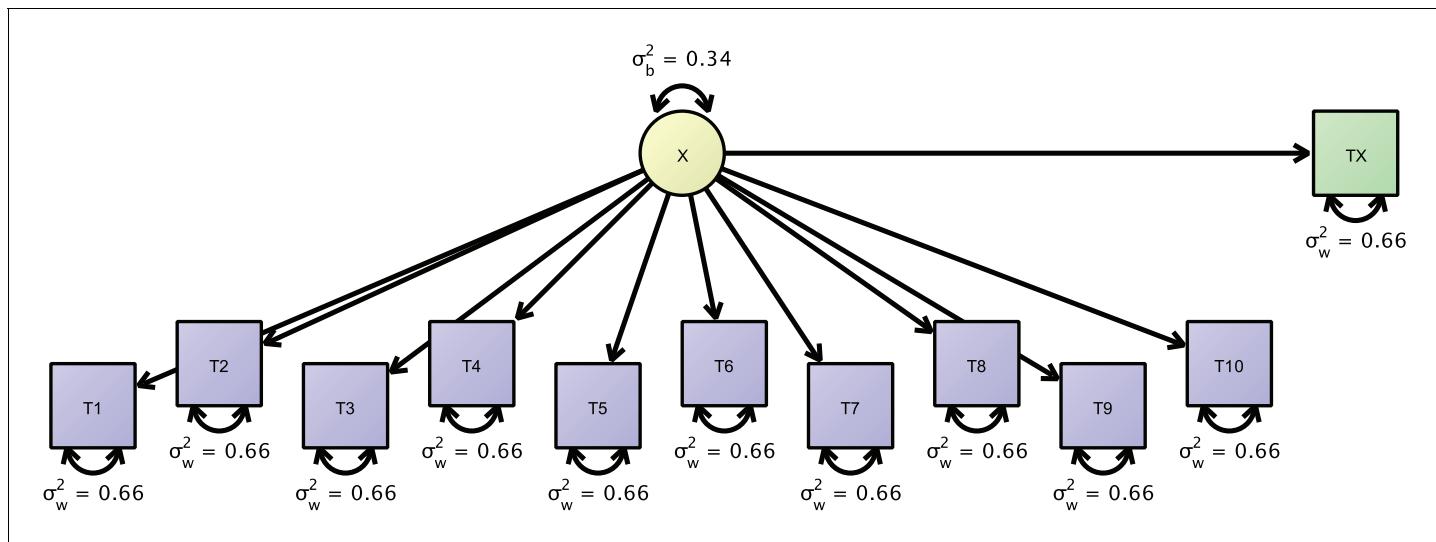


Figure 6. Path-diagram of a joint cross-sectional and longitudinal model to estimate link-wise resting state functional connectivity. The latent variable X (yellow) represents the outcome of interest (e.g., a particular link-wise connectivity coefficient) and is longitudinally measured by 10 measures T1 to T10 (blue rectangles). A second cross-sectional measurement, Tx (green rectangle), augments the estimation of the between-person differences. The respective data set is organized with a mutually missing data scheme, that is, all cells of the longitudinal measurements are missing for cross-sectional data rows and vice versa.

DOI: <https://doi.org/10.7554/eLife.35718.009>

is outweighed by noise by a factor of about 4). The average ICC in our analysis is 0.28 and, thus, very much in line with the original analysis. Compared to [Pannunzi et al., 2018](#), we find a compressed range of ICCs from 0.0 to 0.55 and second the claim that rsFC obtained from 5 min scans performs poorly as a marker for individual subjects (also see [Gordon et al., 2017](#)).

Discussion

When the true scores are changing: Extending ICED to growth curve modeling

So far, we have assumed that the construct of interest does not change over time. Thus, any change between repeated measures was assumed due to unsystematic variability, that is, noise. But what if the construct of interest varies over time? For example, had we modeled all fifty measurements from the day2day study that spanned roughly six month, we would have confounded reliability and lack of stability. it is very likely that the difference between repeated measures in the beginning and at the end of the study represent a mixture of measurement error and true within-person short-term variability, long-term change, or both (also see [Nesselroade, 1991](#)). When assessing reliability over repeated measures in practice, one seeks avoiding this problem by reducing the interval between measurements. At the same time, one is interested in independent measurements, and the degree of dependence may increase with shorter time spans as the chance of item-specific or construct-general temporal effects that may affect multiple measurements may artificially increase the reliability estimate. If, however, measurements are numerous or if the reliability estimate must be obtained from an existing study with a considerable time lag between measurements, it is likely that true change in the construct is present, and that persons differ regarding its magnitude, direction, or both. If substantive change is not accounted for, reliability estimates are biased towards lower values ([Brandmaier et al., 2018](#)). The resulting biased measure may still be useful when interpreted as a stability coefficient, while keeping in mind that instability may be caused by change as well as imprecise measurement. What is, however, the best strategy when we wish to know whether true scores have changed?

Elsewhere, we have applied the logic presented here to linear latent growth curve models ([Brandmaier et al., 2015; Brandmaier et al., 2018; von Oertzen and Brandmaier, 2013](#)). Effective

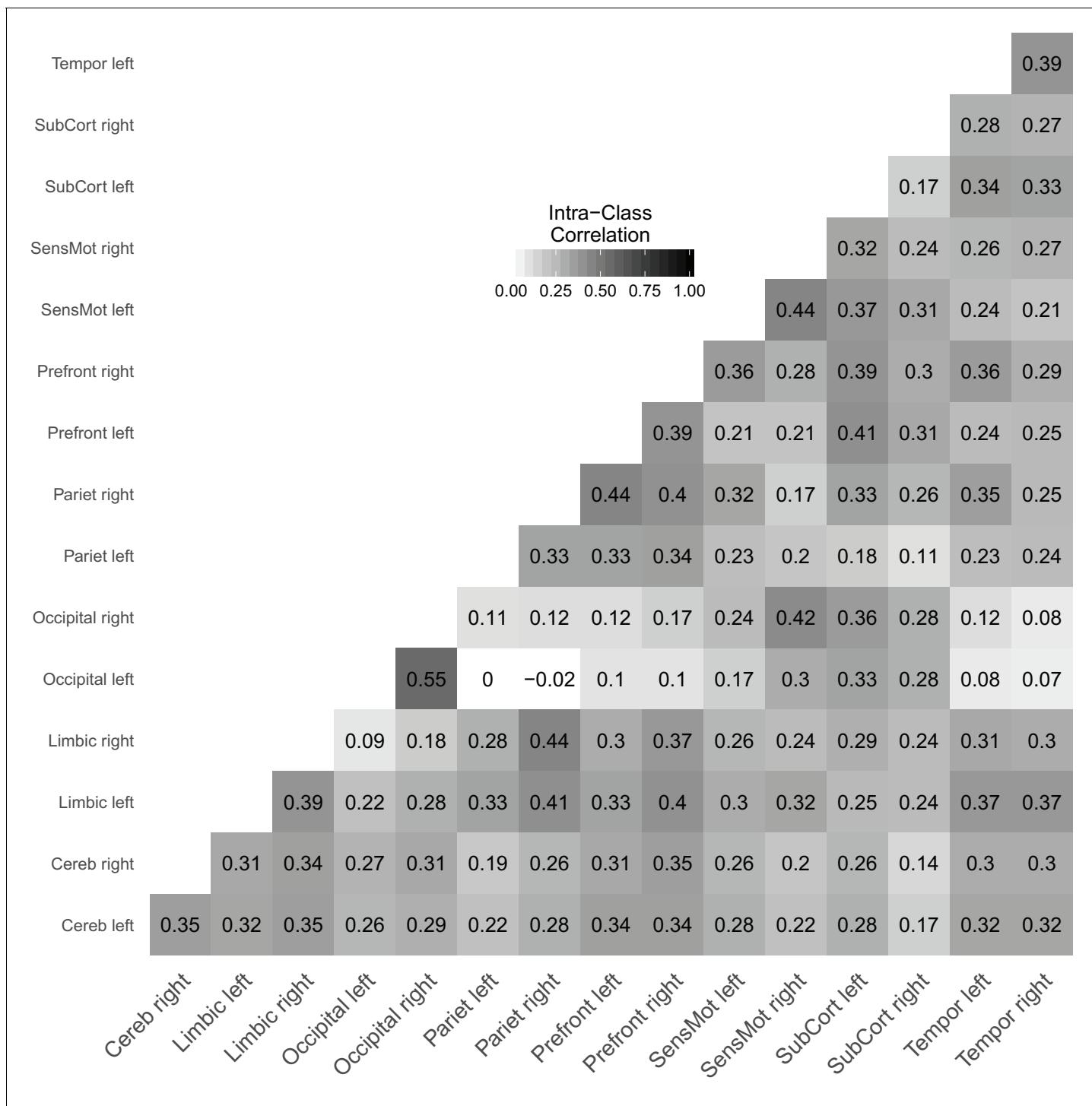


Figure 7. Link-wise reliability based on combined cross-sectional and longitudinal samples from the day2day study.

DOI: <https://doi.org/10.7554/eLife.35718.010>

error of the change component (or, slope) in a latent growth curve model reflects the precision with which a growth curve model can measure between-person differences in change. By scaling the magnitude of individual differences in change (i.e., between-person variance in slope) with effective error, we obtain effective curve reliability (ECR; **Brandmaier et al., 2015**). Major components of effective error for individual differences in change are the number of measurement occasions, the

temporal arrangement of measurement occasions, the total study time span, and instrument reliability. We have shown that effective error, reliability, and statistical power are all potentially useful measures that quantify the sensitivity of a longitudinal design, or any repeated measures design (for example, multiple sessions within a day), to measure individual differences in change (*Brandmaier et al., 2015; Brandmaier et al., 2018*). All these measures may be used for *a priori* design optimization. Such optimization entails either trading-off multiple design factors against each other, while keeping power constant, or changing power as a function of the various design factors and treating them as important measures to communicate reliability of change beyond cross-sectional reliability.

Intra-class effect decomposition of group differences and interactions among error sources

In Section 4, we have discussed a research design that is optimal in factorizing the total error variance into three orthogonal error components of day-specific, session-specific, and unspecific residual variance. Optimality referred to a design that comprises the smallests number of measurements necessary to identify the sought-after error components. However, the ICED framework easily generalizes to more complex designs. For example, with a greater number of sessions, it would be possible to identify additional sources of error, such as experimenter-specific or site-specific errors. We can think of this framework as a variance decomposition approach just as in regular analysis of variance (see *Noble et al., 2017*), with the only difference that we are not interested in the sources of true score variance with the residuals set aside but rather in the decomposing the error score variance.

In the example reported in Section 4, we only examined main effects of day and session. Note, however, that the ICED method also can handle interaction effects. For example, we may be interested if there is an interaction of day and session, that is, if it matters on which day repositioning happened. To test this interaction, one can easily add a second group to the design presented earlier, with two sessions at Day one and one session at Day two (i.e., the mirror image of the current design, in which there is one session at Day one and two sessions at Day two). In this model, we could estimate the ICED components separately for each group. To test a potential interaction, we would state a null hypothesis of no differences in error variance across groups for the session effect. This is a null model of no interaction between session and order-of-day. Explicitly testing the initial model against the restricted null model yields a χ^2 significance test of the interaction. Now, imposing this equality constraint on the session effect across groups would effectively test for the presence of reliable session by day interaction (e.g., does it make a difference whether repositioning within a day takes place at Day one or Day two). One could also conduct the same study with different groups, such as children, older adults, or patients with a particular disease or condition to evaluate group differences in day and session error contributions.

Summary

In this paper, we have discussed the distinction and complementarity of ICC and CV in gauging reliability of brain imaging measures, a topic that thus far has received only limited attention. Considering the increasing demand for longitudinal and multi-center studies, there is a dire need for properly evaluating reliability and identifying components that contribute to measurement error. ICC and CV, as measures of (relative) precision, or reliability, fundamentally relate information about lasting properties of the participants to the precision with which we can measure this information over repeated assessments under the assumption of no change in the underlying construct. We have shown how the generality of the SEM approach (cf. *McArdle, 1994*) may be leveraged to identify components of error sources and estimate their magnitude in more complex designs in more comprehensive and general ways than achievable with standard ANOVA-based ICC decompositions. The underlying framework for deriving the individual error components as factors of reliability is closely related to Cronbach's generalizability theory (or G-Theory; *Cronbach et al., 1972*), which was recently expressed in a SEM framework (*Vispoel et al., 2018*). Our approach is similar to those approaches but was derived using the power equivalence logic (*von Oertzen, 2010*) to analytically derive effective error and reliability scores in a SEM context. This means that our approach easily generalizes to complex measurement designs beyond standard ANOVA, and that effective error, ICC, ICC_2 can

automatically be derived using [von Oertzen \(2010\)](#) algorithm from any study design rendered as a path diagram or in matrix-based SEM notation.

As noted at the beginning of our article, ICC and CV represent two perspectives on reliability that correspond to a fundamental divide of approaches to the understanding of human behavior: the experimental and the correlational (individual differences), each coming with its own notion of reliability ([Cronbach, 1957](#); [Hedge et al., 2018](#)). In experimental settings, reliable effects are usually those that are observed on average, that is, assumed to exist in most individuals. To facilitate detection of such effects, the within-person variability must be low in relation to the average effect. The experimental approach is therefore compatible with the CV perspective. In individual difference approaches, reliable effects distinguish well between persons, which is only true if the within-person variability is low in relation to the between-person variability. The two notions of reliability are associated with competing goals; hence, it is not surprising that robust experimental effects often do not translate into reliable individual differences ([Hedge et al., 2018](#)).

In addition to ICC and CV, other reliability indices have been reported. When researchers compare the similarity of sets, as in gauging the overlap of voxels identified in two repeated analyses of the same subject, the Sørensen–Dice similarity coefficient (or, Dice coefficient; [Dice, 1945](#); [Sørensen, 1948](#)) is often used. Since we are focusing on the reliability of derived continuous indices (e.g., total gray matter volume, fractional anisotropy or indices of myelin water fraction in a region of interest, or link-wise resting state functional connectivity), we did not consider the Dice coefficient here. Others have used the Pearson product moment correlation coefficient, r , to quantify the consistency of test scores across repeated assessments. The linear correlation is a poor choice for reliability assessment because due to its invariance to linear transformation, it is insensitive to mean changes ([Bartko, 1966](#)). Moreover, it is limited to two-occasion data. Therefore, we have also not considered Pearson's r here.

Outlook

Effective error variance partitioning as described above can be useful for communicating absolute precision of measurement, on its own and complimentarily with reliability. Importantly, one needs to specify what kind of reliability is being sought: reliability with respect to an anchoring point (e.g., the scale's zero) or with respect to the heterogeneity in the population. It needs to be emphasized that ICC can only be large if there are individual differences across persons in the measure of interest. Critics of ICC-based approaches to estimating reliability have argued that this method confounds group heterogeneity in the outcome of interest and measurement precision, and therefore must 'be perceived as an extremely misleading criterion for judging the measurement qualities of an instrument.' ([Willett, 1989](#), p. 595). We strongly disagree with this narrow view of measurement quality. In the proverbial sense, 'one man's trash is another man's treasure,' and what some may view as a 'confound,' is for others a virtue of the measure in as much as it determines the capability of detecting heterogeneity in the population. However, the ICC may reveal nothing about the trial-to-trial differences expressed as deviations in the actual unit of measurement; those are better represented by the within-person standard deviation or standardized versions of it. We maintain that ICC is the appropriate measure of reliability when assessing diagnostic instruments and especially while focusing on individual differences.

In this article we introduced ICED as a variance-partitioning framework to quantify the contributions of various measurement context characteristics to unreliability. ICED allows researchers to (1) identify error components; (2) draw inferences about their statistical significance and effect size; and (3) inform the design of future studies.

Given the remarkable pace of progress in human brain imaging, researchers often will be interested in the (yet unknown) reliability of a new neuroimaging measure. Whether this reliability is sufficient can roughly be decided using thresholds, which essentially are a matter of consensus and conventions. For example, reliability larger than 0.9 is often regarded as excellent, as it implies a signal to noise ratio of 10:1. However, there may be good reasons to adopt less conservative thresholds (e.g., [Cicchetti and Sparrow, 1981](#)). In addition, using ICED, researchers can go beyond a summary index of ICC and instead report the magnitudes of individual variance components that contribute to lowering the overall ICC. These different components may differ in their methodological and practical implications. Often, researchers will be interested in using inferential statistics to test whether each of the individual variance components differs from zero and, maybe, whether the components

differ from each other. Finally, the results of these analyses can guide researchers in their subsequent attempts to improve measurement reliability. For instance, using ICED, researchers may discover that a hitherto overlooked but remediable source of error greatly contributes to unreliability, and work on improving the measurement properties influencing this component. Also, researchers may ask what combinations of measurements are needed to attain a target reliability (**Noble et al., 2017**) while optimizing an external criterion such as minimizing costs or participant burden (**Brandmaier et al., 2015**).

To conclude, we hope that the tools summarized under ICED will be applied in human brain imaging studies to index overall reliability, and to identify and quantify multi-source contributions to measurement error. We are confident that the use of ICED will help researcher to develop more reliable measures, which are a prerequisite for more valid studies.

Acknowledgements

We thank Muzamil Arshad and Jeffrey A Stanley from the Department of Psychiatry and Behavioral Neuroscience, School of Medicine, Wayne State University, Detroit, Michigan, for providing the raw data on Myelin Water Fraction measurements. This work was supported by European Union's Horizon 2020 research and innovation programme under grant agreement No. 732592: 'Healthy minds from 0–100 years: optimizing the use of European brain imaging cohorts ('Lifebrain') to AB, SK and UL, and by NIH grant R01-AG011230 to NR.

Additional information

Funding

Funder	Grant reference number	Author
Horizon 2020 Framework Programme	732592	Andreas M Brandmaier Simone Kühn Ulman Lindenberger
Max-Planck-Gesellschaft		Andreas M Brandmaier Elisabeth Wenger Nils C Bodammer Naftali Raz Ulman Lindenberger Simone Kühn
National Institutes of Health	R01-AG011230	Naftali Raz

Open-access funding. The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication

Author contributions

Andreas M Brandmaier, Conceptualization, Data curation, Formal analysis, Visualization, Methodology, Writing original draft, Writing—review and editing; Elisabeth Wenger, Nils C Bodammer, Conceptualization, Methodology, Writing—review and editing; Simone Kühn, Resources, Data curation, Investigation, Writing—review and editing; Naftali Raz, Conceptualization, Resources, Data curation, Supervision, Methodology, Writing—review and editing; Ulman Lindenberger, Conceptualization, Supervision, Methodology, Writing—review and editing

Author ORCIDs

Andreas M Brandmaier  <http://orcid.org/0000-0001-8765-6982>

Naftali Raz  <http://orcid.org/0000-0002-5080-2138>

Ulman Lindenberger  <http://orcid.org/0000-0001-8428-6453>

Decision letter and Author response

Decision letter <https://doi.org/10.7554/eLife.35718.016>

Author response <https://doi.org/10.7554/eLife.35718.017>

Additional files

Data availability

The dataset on myelin water fraction measurements is freely available at <https://osf.io/t68my/> and the link-wise resting state functional connectivity data are available at <https://osf.io/8n24x/>.

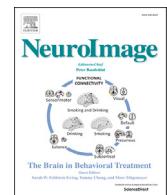
The following previously published datasets were used:

Author(s)	Year	Dataset title	Dataset URL	Database, license, and accessibility information
Pannunzi M, Hindriks R, Bettinardi R, Wenger E, Lisofsky N, Mårtensson J, Butler O, Filevich E, Becker M, Lochstet M, Lindenberger U, Kühn S, Deco G	2018	Resting-state fMRI correlations: from link-wise unreliability to whole brain stability	https://osf.io/8n24x/	Publicly available at Open Science Framework
Arshad M, Stanley J A, Raz N	2018	Reliability of Myelin Water Fraction in ALIC	https://osf.io/t68my/	Publicly available at Open Science Framework

References

- Arshad M, Stanley JA, Raz N. 2017. Test-retest reliability and concurrent validity of in vivo myelin content indices: Myelin water fraction and calibrated T₁ w/T₂ w image ratio. *Human Brain Mapping* **38**:1780–1790. DOI: <https://doi.org/10.1002/hbm.23481>, PMID: 28009069
- Arshad M, Stanley JA, Raz N. 2018. Reliability of myelin water fraction in ALIC. OSF. <https://osf.io/t68my/files/>
- Bartko JJ. 1966. The intraclass correlation coefficient as a measure of reliability. *Psychological Reports* **19**:3–11. DOI: <https://doi.org/10.2466/pr0.1966.19.1.3>, PMID: 5942109
- Bliese PD. 2000. Within-group agreement, non-independence, and reliability: Implications for data aggregation and analysis. In: *Multilevel Theory, Research, Andmethods in Organizations: Foundations, Extensions, and New Directions*. San Francisco, CA: Jossey-Bass. p. 349–381.
- Boker SM, McArdle JJ, Neale M. 2002. An algorithm for the hierarchical organization of path diagrams and calculation of components of expected covariance. *Structural Equation Modeling: A Multidisciplinary Journal* **9**: 174–194. DOI: https://doi.org/10.1207/S15328007SEM0902_2
- Brandmaier AM, von Oertzen T, Ghisletta P, Hertzog C, Lindenberger U. 2015. LIFESPAÑA: a tool for the computer-aided design of longitudinal studies. *Frontiers in Psychology* **6**. DOI: <https://doi.org/10.3389/fpsyg.2015.00272>, PMID: 25852596
- Brandmaier AM, von Oertzen T, Ghisletta P, Lindenberger U, Hertzog C. 2018. Precision, reliability, and effect size of slope variance in latent growth curve models: implications for statistical power analysis. *Frontiers in Psychology* **9**:294. DOI: <https://doi.org/10.3389/fpsyg.2018.00294>, PMID: 29755377
- Cicchetti DV, Sparrow SA. 1981. Developing criteria for establishing interrater reliability of specific items: applications to assessment of adaptive behavior. *American Journal of Mental Deficiency* **86**:127–137. PMID: 7315877
- Craddock RC, Holtzheimer PE, Hu XP, Mayberg HS. 2009. Disease state prediction from resting state functional connectivity. *Magnetic Resonance in Medicine* **62**:1619–1628. DOI: <https://doi.org/10.1002/mrm.22159>, PMID: 19859933
- Cronbach LJ, Gleser GC, Nanda H, Rajaratnam N. 1972. *The Dependability of Behavioral Measurements: Theory of Generalizability for Scores and Profiles*. New York: John Wiley.
- Cronbach LJ. 1957. The two disciplines of scientific psychology. *American Psychologist* **12**:671–684. DOI: <https://doi.org/10.1037/h0043943>
- de Schoot RV, Hoijtink H, Jan-Willem R. 2011. Moving beyond traditional null hypothesis testing: evaluating expectations directly. *Frontiers in Psychology* **2**:24. DOI: <https://doi.org/10.3389/fpsyg.2011.00024>, PMID: 21713172
- Dice LR. 1945. Measures of the amount of ecological association between species. *Ecology* **26**:297–302. DOI: <https://doi.org/10.2307/1932409>
- Dosenbach NU, Nardos B, Cohen AL, Fair DA, Power JD, Church JA, Nelson SM, Wig GS, Vogel AC, Lessov-Schlaggar CN, Barnes KA, Dubis JW, Feczkó E, Coalson RS, Pruett JR, Barch DM, Petersen SE, Schlaggar BL. 2010. Prediction of individual brain maturity using fMRI. *Science* **329**:1358–1361. DOI: <https://doi.org/10.1126/science.1194144>, PMID: 20829489
- Filevich E, Lisofsky N, Becker M, Butler O, Lochstet M, Martensson J, Wenger E, Lindenberger U, Kühn S. 2017. Day2day: investigating daily variability of magnetic resonance imaging measures over half a year. *BMC Neuroscience* **18**:65. DOI: <https://doi.org/10.1186/s12868-017-0383-y>, PMID: 28836958

- Finkbeiner C. 1979. Estimation for the multiple factor model when data are missing. *Psychometrika* **44**:409–420. DOI: <https://doi.org/10.1007/BF02296204>
- Garrett DD, Samanez-Larkin GR, MacDonald SW, Lindenberger U, McIntosh AR, Grady CL. 2013. Moment-to-moment brain signal variability: a next frontier in human brain mapping? *Neuroscience & Biobehavioral Reviews* **37**:610–624. DOI: <https://doi.org/10.1016/j.neubiorev.2013.02.015>, PMID: 23458776
- Gee DG, McEwen SC, Forsyth JK, Haut KM, Bearden CE, Addington J, Goodyear B, Cadenhead KS, Mirzakhanian H, Cornblatt BA, Olvet D, Mathalon DH, McGlashan TH, Perkins DO, Belger A, Seidman LJ, Thermenos H, Tsuang MT, van Erp TG, Walker EF, et al. 2015. Reliability of an fMRI paradigm for emotional processing in a multisite longitudinal study. *Human Brain Mapping* **36**:2558–2579. DOI: <https://doi.org/10.1002/hbm.22791>, PMID: 25821147
- Gordon EM, Laumann TO, Gilmore AW, Newbold DJ, Greene DJ, Berg JJ, Ortega M, Hoyt-Drazen C, Gratton C, Sun H, Hampton JM, Coalson RS, Nguyen AL, McDermott KB, Shimony JS, Snyder AZ, Schlaggar BL, Petersen SE, Nelson SM, Dosenbach NUF. 2017. Precision functional mapping of individual human brains. *Neuron* **95**: 791–807. DOI: <https://doi.org/10.1016/j.neuron.2017.07.011>, PMID: 28757305
- Hedge C, Powell G, Sumner P. 2018. The reliability paradox: why robust cognitive tasks do not produce reliable individual differences. *Behavior Research Methods* **50**:1166–1186. DOI: <https://doi.org/10.3758/s13428-017-0935-1>
- Lerch JP, van der Kouwe AJ, Raznahan A, Paus T, Johansen-Berg H, Miller KL, Smith SM, Fischl B, Sotiroopoulos SN. 2017. Studying neuroanatomy using MRI. *Nature Neuroscience* **20**:314–326. DOI: <https://doi.org/10.1038/nn.4501>, PMID: 28230838
- McArdle JJ. 1980. Causal modeling applied to psychonomic systems simulation. *Behavior Research Methods & Instrumentation* **12**:193–209. DOI: <https://doi.org/10.3758/BF03201598>
- McArdle JJ. 1994. Structural factor analysis experiments with incomplete data. *Multivariate Behavioral Research* **29**:409–454. DOI: https://doi.org/10.1207/s15327906mbr2904_5, PMID: 26745236
- Morey RA, Selgrade ES, Wagner HR, Huettel SA, Wang L, McCarthy G. 2010. Scan-rescan reliability of subcortical brain volumes derived from automated segmentation. *Human Brain Mapping* **31**:1751–1762. DOI: <https://doi.org/10.1002/hbm.20973>, PMID: 20162602
- Nesselroade JR. 1991. *The Warp and Woof of the Developmental Fabric* Hillsdale. Hillsdale, NJ: Lawrence Erlbaum. p. 213–240.
- Noble S, Spann MN, Tokoglu F, Shen X, Constable RT, Scheinost D. 2017. Influences on the Test-Retest reliability of functional connectivity MRI and its relationship with behavioral utility. *Cerebral Cortex* **27**:5415–5429. DOI: <https://doi.org/10.1093/cercor/bhw230>, PMID: 28968754
- Pannunzi M, Hindriks R, Bettinardi RG, Wenger E, Lisofsky N, Martensson J, Butler O, Filevich E, Becker M, Lochstet M, Lindenberger U, Kühn S, Deco G. 2018. Corrigendum to “Resting-state fMRI correlations: From link-wise unreliability to whole brain stability”. *NeuroImage* **174**:599–604. DOI: <https://doi.org/10.1016/j.neuroimage.2017.12.028>
- Raykov T, Marcoulides GA, Patelis T. 2015. The importance of the assumption of uncorrelated errors in psychometric theory. *Educational and Psychological Measurement* **75**:634–647. DOI: <https://doi.org/10.1177/0013164414548217>, PMID: 29795836
- Rosseel Y. 2012. lavaan : an R package for structural equation modeling. *Journal of Statistical Software* **48**:1–36. DOI: <https://doi.org/10.18637/jss.v048.i02>
- Sørensen T. 1948. A method of establishing groups of equal amplitude in plant sociology based on similarity of species and its application to analyses of the vegetation on danish commons. *Biologiske Skrifter* **5**:1–34.
- Vispoel WP, Morris CA, Kilinc M. 2018. Applications of generalizability theory and their relations to classical test theory and structural equation modeling. *Psychological Methods* **23**:1–26. DOI: <https://doi.org/10.1037/met0000107>, PMID: 28114776
- von Oertzen T, Brandmaier AM, Tsang S. 2015. Structural equation modeling with onyx. *Structural Equation Modeling: A Multidisciplinary Journal* **22**:148–161. DOI: <https://doi.org/10.1080/10705511.2014.935842>
- von Oertzen T, Brandmaier AM. 2013. Optimal study design with identical power: an application of power equivalence to latent growth curve models. *Psychology and Aging* **28**:414–428. DOI: <https://doi.org/10.1037/a0031844>, PMID: 23586357
- von Oertzen T, Hertzog C, Lindenberger U, Ghisletta P. 2010. The effect of multiple indicators on the power to detect inter-individual differences in change. *British Journal of Mathematical and Statistical Psychology* **63**:627–646. DOI: <https://doi.org/10.1348/000711010X486633>, PMID: 20211053
- von Oertzen T. 2010. Power equivalence in structural equation modelling. *British Journal of Mathematical and Statistical Psychology* **63**:257–272. DOI: <https://doi.org/10.1348/000711009X441021>, PMID: 19527562
- Willett JB. 1989. Some results on reliability for the longitudinal measurement of change: implications for the design of studies of individual growth. *Educational and Psychological Measurement* **49**:587–602. DOI: <https://doi.org/10.1177/001316448904900309>



A novel training-free externally-regulated neurofeedback (ER-NF) system using phase-guided visual stimulation for alpha modulation



Gan Huang^{a,b}, Jia Liu^{a,b}, Linling Li^{a,b}, Li Zhang^{a,b}, Yixuan Zeng^c, Lijie Ren^c, Shiqing Ye^d, Zhiguo Zhang^{a,b,*}

^a School of Biomedical Engineering, Health Science Center, Shenzhen University, Shenzhen, 518060, China

^b Guangdong Provincial Key Laboratory of Biomedical Measurements and Ultrasound Imaging, Shenzhen, 518060, China

^c Department of Neurology, Shenzhen Second People's Hospital, Shenzhen University 1st Affiliated Hospital, Shenzhen, 518029, China

^d School of Information Science and Technology, Fudan University, Shanghai, 200433, China

ARTICLE INFO

Keywords:
 Neurofeedback
 Self-regulation
 Alpha wave
 Training-free
 Visual stimulation

ABSTRACT

The efficacy of neurofeedback is a point of great controversy, because a certain proportion of users cannot properly regulate their brain activities and thereby fail to benefit from neurofeedback. To address the neurofeedback inefficacy problem, the present study is aimed to design and implement a new neurofeedback system that can more effectively and consistently regulate users' brain activities than the conventional way of training users to voluntarily regulate brain activities. The new neurofeedback system delivers external visual stimuli continuously at a specific alpha phase, which is real-time decoded from ongoing alpha wave, to regulate the alpha wave. Experimental results show that the proposed training-free externally-regulated neurofeedback (ER-NF) system can achieve consistent (effective in almost all sessions for almost all users), flexible (either increasing or decreasing peak alpha frequency and alpha power), and immediate (taking or losing effect immediately after stimulation is on or off) modulation effects on alpha wave. Therefore, the ER-NF system holds great potential to be able to more reliably and flexibly modulate cognition and behavior.

1. Introduction

Neurofeedback is aimed to guide people to voluntarily regulate their brain activities to desired patterns by measuring and showing their relevant brain activity patterns (Hammond et al., 2007). The most commonly regulated brain activity patterns used in neurofeedback are electroencephalographic (EEG) rhythms. Jasper and Shagass (Jasper and Shagass, 1941) first showed that human subjects could be instructed to voluntarily regulate alpha wave, which laid the foundation for neurofeedback. Nowadays, neurofeedback has been used as a therapeutic intervention for the treatment of a range of brain diseases and disorders, such as attention-deficit/hyperactivity disorder (ADHD) (Albrecht et al., 2015), epilepsy (Tan et al., 2009), stroke (Mihara et al., 2013), autistic spectrum disorder (ASD) (Coben et al., 2010), and emotional disorders (Linden et al., 2012).

However, the efficacy of neurofeedback is still a point of great controversy. For example, although many empirical studies supported neurofeedback's efficacy in the treatment of ADHD (Fuchs et al., 2003; Monastra et al., 2002), a few well-designed and -controlled studies

reported either absent or reduced effects (Moriyama et al., 2012; Holtmann et al., 2014). Even clinical effects of neurofeedback do exist, it still remains unknown whether the effects are caused by neurofeedback itself or just placebo effects mediated by expectancy (Lofthouse et al., 2012). Neurofeedback inefficacy is also manifested by the fact that a significant proportion of users cannot benefit from neurofeedback (Alkoby et al., 2018). For example, Doehnert et al. reported that about half of their subjects did not succeed in regulating brain activity in their neurofeedback training with ADHD patients (Doehnert et al., 2008). In another study, Lubar et al. found that about 40% of their subjects were not able to regulate EEG to the desired pattern even after 40 sessions of neurofeedback training (Lubar et al., 1995). Neurofeedback inefficacy could be attributed to many factors, such as psychological characteristics and physiological states (Alkoby et al., 2018). One well-recognized reason underlying neurofeedback inefficacy is the great difficulty for users (at least, a certain proportion of users) to master the skills necessary to self-regulate their certain brain activities to match desired patterns. As a consequence, those people who are unable to regulate their brain activities fail to achieve any positive effects. Because of the limitation of self-regulation, a new regulation technique that can modulate brain

* Corresponding author. School of Biomedical Engineering, Health Science Center, Shenzhen University, Shenzhen, 518060, China.
 E-mail address: zgzhang@szu.edu.cn (Z. Zhang).

Abbreviations

ER-NF	Externally-Regulated Neurofeedback (the proposed neurofeedback protocol)
SR-NF	Self-Regulated Neurofeedback (the traditional neurofeedback protocol)
REST	resting-state EEG with eyes-open
VEP	visual-evoked potentials
SSVEP	steady-state visual-evoked potentials
F_{rest}	the peak frequency of alpha wave (8–12 Hz) in REST
I_{LED}	the light intensity of LED used in ER-NF
P_{rest}	the peak power of alpha wave at F_{rest} in REST
F_{mod}	the main frequency of the modulation function (alpha power against phase) in ER-NF

activities without users' active participation is highly desired in developing new neurofeedback protocols.

In the present study, we proposed a new neurofeedback protocol that uses external visual stimulation, which is generated based on real-time decoded phases of alpha wave, to regulate the frequency and power of alpha wave. By using external sensory stimulation to regulate users' brain activities, users' active participation is not necessary and they do not need to master any learning skill or experience of self-regulation. Therefore, the new neurofeedback protocol based on such an external regulation of EEG activities is expected to achieve universal and consistent regulation effects. Hereinafter, the new neurofeedback protocol is referred to as externally-regulated neurofeedback (ER-NF), while the traditional neurofeedback based on self-regulation is referred to as Self-Regulated neurofeedback (SR-NF). The key novelty of the proposed ER-NF protocol is that, the delivery time of visual stimuli used to modulate alpha wave is determined by the phases of ongoing alpha wave. It makes the new neurofeedback protocol largely different from conventional visual evoked potential (VEP) protocols using prespecified delivery time. The phase-guided visual stimulation can regulate alpha wave in a more adaptive and flexible manner, and its principle and advantages are explained as follows.

The idea of using phase-guided visual stimulation to regulate EEG comes from the observation that, the timing of stimulation (which corresponds to the phase of certain EEG rhythms when the stimulation is delivered) has a huge impact on brain responses. For example, visual stimuli delivered at different phases of alpha wave evoke different alpha dynamic responses (McSayers and Beagley, 1974; Trimble et al., 1975; Jervis et al., 1983). In another word, the latency and amplitude of evoked

potentials are sensitive to the phase of the alpha wave at the stimulation time. This observation inspires us to continuously deliver visual stimuli at a specific alpha phase to modulate the alpha wave to expected dynamic behaviors, which is the basis of the proposed new ER-NF protocol. The principle of phase-guided visual stimulation is illustrated in Fig. 1, where we simplify the alpha oscillation behavior as the motion trajectory of a simple pendulum. Without damping, the simple pendulum system will execute a simple harmonic motion with the frequency and amplitude unchanged (Fig. 1a). However, if we exert a force on the single pendulum at a specific phase, the amplitude of the single pendulum would increase or decrease, depending on the phase of pendulum at which the force is exerted (Fig. 1b and c). Correspondingly, the system frequency would also be changed. Hence, if we can keep delivering a series of visual stimuli at one specific alpha phase, we could change the pattern of the alpha wave (i.e., the amplitude and the frequency of the alpha wave).

There are still a number of difficulties to be addressed in the design and implementation of such an ER-NF system using phase-guided visual stimuli. It is vital to precisely real-time decode the alpha phase for the determination of delivery time of visual stimuli, and it is important to optimize stimulation parameters for each user to maximize the regulation effects. The present study addressed above difficulties and we showed that a proof-of-concept real-time ER-NF system could effectively increase or decrease the power and the frequency of alpha wave. More importantly, the regulation effects of the ER-NF system can be reliably observed in almost all experiments and from almost all users, even users were not trained to master any regulation skills. Hence, this new training-free ER-NF protocol holds great potential to be able to more effectively and consistently modulate cognition and behavior.

2. Methods

2.1. System design

The schema for the proposed ER-NF system is shown in Fig. 2. The system consists of four modules: (1) EEG recording, (2) phase decoding, (3) stimulation sequence generation, (4) visual stimulation. Raw EEG signals were recorded and then the phase of alpha wave was estimated and used to guide the generation of visual stimulation sequence (i.e., the exact time to deliver visual stimuli). Visual stimuli were delivered by LED to provide feedback to users. Because the sequence of visual stimuli was generated based on the alpha phase and then, in turn, modulated the alpha wave, a closed-loop control of the alpha wave was formed. The details of the four modules are introduced as follows.

2.1.1. Module 1: EEG recording

Raw EEG was recorded by a BrainAmp system (Brain Products GmbH,

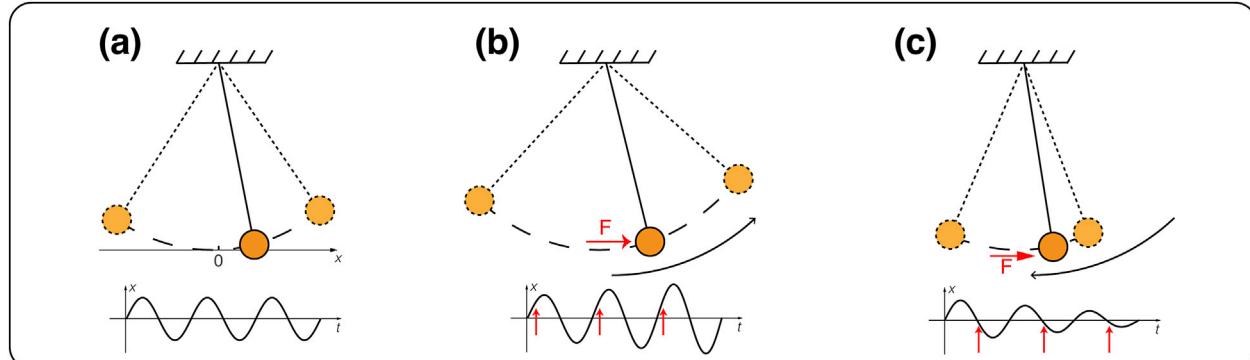


Fig. 1. The principle of the proposed Externally-Regulated Neurofeedback (ER-NF) protocol. (a) The alpha oscillation behavior is simplified as the motion trajectory of a simple pendulum without damping. The motion trajectory of the simple pendulum is a sinusoidal function with fixed frequency and amplitude. The amplitude indicates horizontal displacement of the ball. (b) If a force (the red arrow F in the figure) is exerted at a phase where the pendulum moves in the same direction as the force, the amplitude of the pendulum would be increased. (c) On the contrary, if the force is exerted at the phase where the pendulum moves in the reverse direction as the force, the amplitude of the pendulum would be decreased.

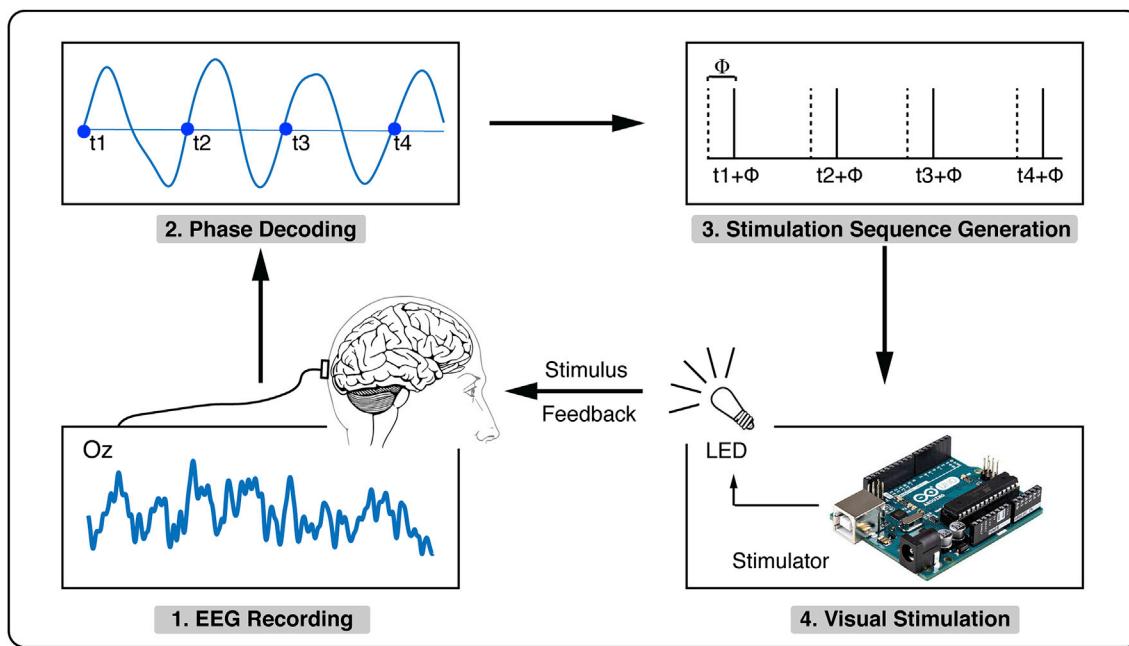


Fig. 2. The schema of the Externally-Regulated Neurofeedback (ER-NF) system.

Germany) from Oz (referenced to FCz) with a sampling rate of 5000 Hz. The impedance for both Oz and the reference electrodes were kept lower than 5 kΩ in the experiments.

2.1.2. Module 2: phase decoding

This module was developed in C++ using BrainAmp SDK. Because fast and accurate estimation of the alpha phase is of key importance to the modulation effect (i.e., how and to what extent the alpha wave is modulated) of ER-NF, we used a bandpass filter and a zero-cross point detection method for phase estimation. First, raw EEG was bandpass filtered online by a second-order Butterworth filter with a bandwidth of 2 Hz to separate alpha wave. The center frequency of the bandpass filter was user-dependent and it was equal to the peak frequency of the user's alpha wave at resting state with eyes-open. Hereinafter we used F_{rest} to denote the peak frequency of alpha wave at rest and it was estimated as the frequency with the maximum power in the range of 8–12 Hz. Note that the Butterworth filter introduced a phase delay of around 240 ms, which may influence modulation effect of ER-NF (see Discussion for details). Next, we need to detect one certain alpha phase for the generation of visual stimuli. Here, the alpha phase at $\varphi = 3\pi/2$ was estimated as the upward zero-crossing of the filtered signal. The zero-crossing method was adopted here because its low complexity and effective phase estimation under noise were suitable for online processing. Note that stimuli delivered at different phases would lead to different alpha dynamic behaviors (as explained in the next paragraph), which could counteract with each other so that the overall modulation effect is reduced or even disappears. Hence, we need to deliver visual stimuli around the same alpha phase, which was generated based on the detected phase of $3\pi/2$, to guarantee modulation effects on alpha wave.

2.1.3. Module 3: stimulation sequence generation

Further, we need to generate a sequence of visual stimuli to modulate the alpha wave. The modulation effect depends on the alpha phase when a visual stimulus is perceived, as illustrated in Fig. 1. Thus, we examined the modulation effects of visual stimuli delivered at different alpha phases, which was achieved by adding a time lag Φ after the decoded phase of $3\pi/2$. Twenty different values of time lag Φ (ranging from 0 to 190 ms with a step of 10 ms) were examined and they correspond to 20 phases of alpha wave. Because F_{rest} is around 10 Hz and the corresponding period is close to 100 ms, the range of Φ (0–190 ms) roughly

covers 2 cycles of the alpha wave. The actual delivery time of visual stimuli is around the phase $[3\pi/2 + (\text{system delay} + \Phi) \times 2\pi F_{rest}]$. Since the system delay was a constant (see Discussion for details), we can use the time lag Φ as a phase index. By examining the alpha power at each phase index Φ , we had a modulation function describing how alpha power varied with phase. Based on the simple pendulum model, we expect to see that the modulation effect is periodical and the period is the same as F_{rest} .

2.1.4. Module 4: visual stimulation

Visual stimuli were delivered by LED, which was controlled by a microcontroller (Arduino UNO, Arduino) and placed 45 cm away from subjects' eyes. Ten exponentially increased levels of LED intensity (5, 10, 19, 37, 71, 145, 285, 546, 1074, 1998 Lux, denoted as Level 1 to Level 10) were tested to determine the optimal LED intensity for each subject (see Section 2.3 for details about the selection of light intensity). The light intensity was measured by a light meter (TES-1332A, TES). It should be noted that the light intensities tested in this experiment were in general small. At low levels of light intensity (specially for Level 1 and Level 2), a few subjects could not perceive whether the light was on or off. Even at the highest level of light intensity (Level 10), some subjects still had no clear EEG responses, i.e., VEP.

2.2. Experimental design

Twenty-one healthy subjects (5 females and 16 males) aged 18–26 years (mean 23.57 ± 2.11), without a history of epilepsy, participated in the study. To validate the reproducibility of modulation effect, these subjects took part in two experiments, which were almost the same and were arranged in two different weeks within a month (the interval between one subject's two experiments ranged from 5 to 31 days, mean 15.7 ± 7.50). One subject was excluded from further analysis because he did not take part in the second experiment. As a result, we have totally 20 subjects and 40 experiments. In the following, we use the suffix 'a' and 'b' to denote the first and second experiments, respectively. For example, the second experiment of subject 02 is labeled as Sub02b. The experiments were in accordance with the Declaration of Helsinki. Ethical approval of the study was sought and obtained from the Bioethics Committee, Shenzhen University Health Science Center. Each subject was given the written informed consent prior to the experiments.

Each experiment consisted of two parts: a calibration part and an evaluation part. In the calibration part, two important parameters, I_{LED} (the light intensity of LED) and F_{rest} (the peak frequency of the alpha wave in the resting state with eyes-open) were respectively obtained from VEP and REST (resting-state EEG with eyes-open). In the evaluation part, we evaluated the modulation effect of the proposed ER-NF system in six consecutive sessions. To compare the modulation effects of different types of visual stimulation, SSVEP (steady-state visual-evoked potentials) was also recorded in users' second experiments. The experimental paradigm is illustrated in Fig. 3 and more details are provided below.

2.2.1. Part 1: calibration

- **VEP:** There were ten VEP sessions in the calibration part. In each session, a subject received continuously flashing visual stimuli via LED with random time intervals ranging from 80 ms to 120 ms, so that VEP had similar stimulation frequency (around 10 Hz) as in ER-NF and SSVEP. Each session lasted 2 min. One level of light intensity was used once in a session in a random order. More details about the light intensity were provided in Section 2.1.
- **REST:** There were four resting state sessions: two with eyes open and two with eyes closed. Four sessions were carried out alternately in turn and each session lasted one minute. Only resting state EEG data with eyes open, referred as REST below, were used in subsequent analyses.

2.2.2. Part 2: evaluation

- **ER-NF:** There were six ER-NF sessions. In each session, ER-NF was performed with 20 different alpha phases (Φ) of the alpha wave in a random order, and visual stimulation at each phase lasted 20 s. Since visual stimuli continuously delivered at different phases could not be discriminated by eyes, both the subjects and the experimenter did not know which phase was performed during the experiment. Hence, the ER-NF experiments were double-blind.
- **SSVEP:** Flashing light at the frequency of 10 Hz was used in SSVEP for 30 s. The I_{LED} was determined in the calibration part of each experiment. The stimulus frequency was fixed as 10 Hz (see Discussion for details).

2.3. Parameter selection

Two important parameters used to generate visual stimulation in ER-NF are F_{rest} and I_{LED} , and they can be estimated from REST and VEP data, respectively. To estimate these two important parameters, an offline analysis was run immediately after the calibration part, in which REST and VEP were recorded. A MATLAB script was written for the offline analysis, which lasted less than 2 min. The subjects took a rest during the time.

To detect the F_{rest} for individualized ER-NF, we located the maximum of the power spectrum of REST data (concatenated from EEG of two REST sessions with eyes open) in the alpha band (8–12 Hz). The Welch's method (with a window of 2 s and 50% overlap) was used for power spectral density estimation. Because the alpha wave was variable within subjects, the detected F_{rest} in two separate experiments of the same subject varied slightly, especially when the alpha wave was weak.

The light intensity I_{LED} is also an important parameter. It should be large enough to produce an evident modulation effect but cannot be too large so that users may feel uncomfortable. From the analogy of the simple pendulum in Fig. 1, we can see that a small force is sufficient to change the system's dynamic behavior while a large force might make the system unstable. Therefore, I_{LED} was selected as the minimum light intensity which can evoke VEP in the present study. In our experiments, I_{LED} was determined as the minimum light intensity which evoked a clear VEP response different from the background ongoing EEG. To improve the consistency of this selection criteria, we first made sure all the experiments were performed by the same operator. Moreover, all the light intensity selection results were double-checked by two people (the operator and another one) in offline analyses, and there was no difference between the offline selection of intensity and the intensity used in the experiments. In case there is no clear VEP response for a subject, the highest light intensity will be selected for this subject. An example was provided in Fig. 4 to illustrate how we determine I_{LED} for Sub01b. Because the intervals between consecutive stimuli, 80–120 ms, was too short, it would be difficult to detect VEP from a single visual stimulus by averaging. Hence, we firstly filtered the raw EEG by 8–12 Hz bandpass filter, and then calculated the mean absolute value of the filtered signal under different levels of light intensity. After baseline correction with the interval -500 – 0 ms, the Hilbert transform was applied on the signal. Finally, the amplitudes of the Hilbert transformed VEP signals are shown in Fig. 4 with ten different light intensities for Sub01b. The amplitudes of VEP (measured as the averaged values in the interval of 50–300 ms) are displayed in Fig. 4. In this example, the light intensity at level 7 was selected as the I_{LED} in ER-NF, because VEP with the light intensity at this level had a clear response different from ongoing EEG while VEP evoked by the light intensity from Level 1 to Level 6 did not.

2.4. Data analyses

2.4.1. Single-subject analyses of modulation effects

To evaluate the modulation effect of ER-NF at a range of alpha phases, the power spectral density of regulated alpha wave was estimated using the Welch's method (the window length is set to be 2 s with 50% overlap). Because ER-NF was expected to modulate the pattern of alpha wave (i.e., the distribution of alpha power over frequency), we mainly checked how the alpha power was modulated at one specific frequency F_{rest} and also examined how the peak alpha frequency was modulated. The power of the alpha wave was estimated as the averaged spectral power over

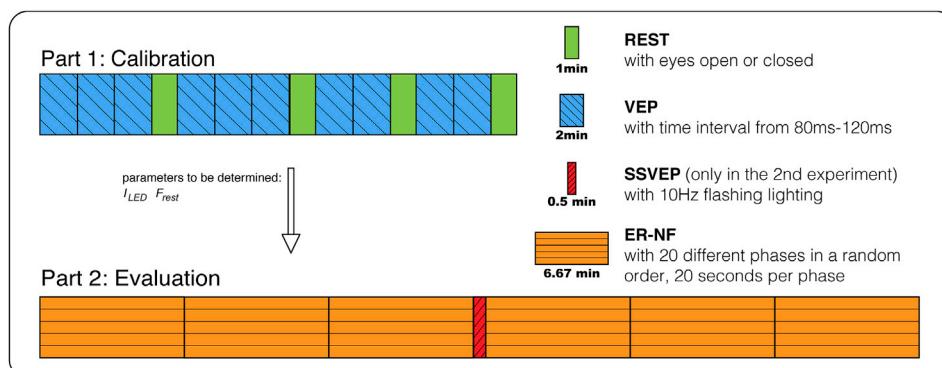


Fig. 3. Experimental paradigm. All subjects took the experiment for two times.

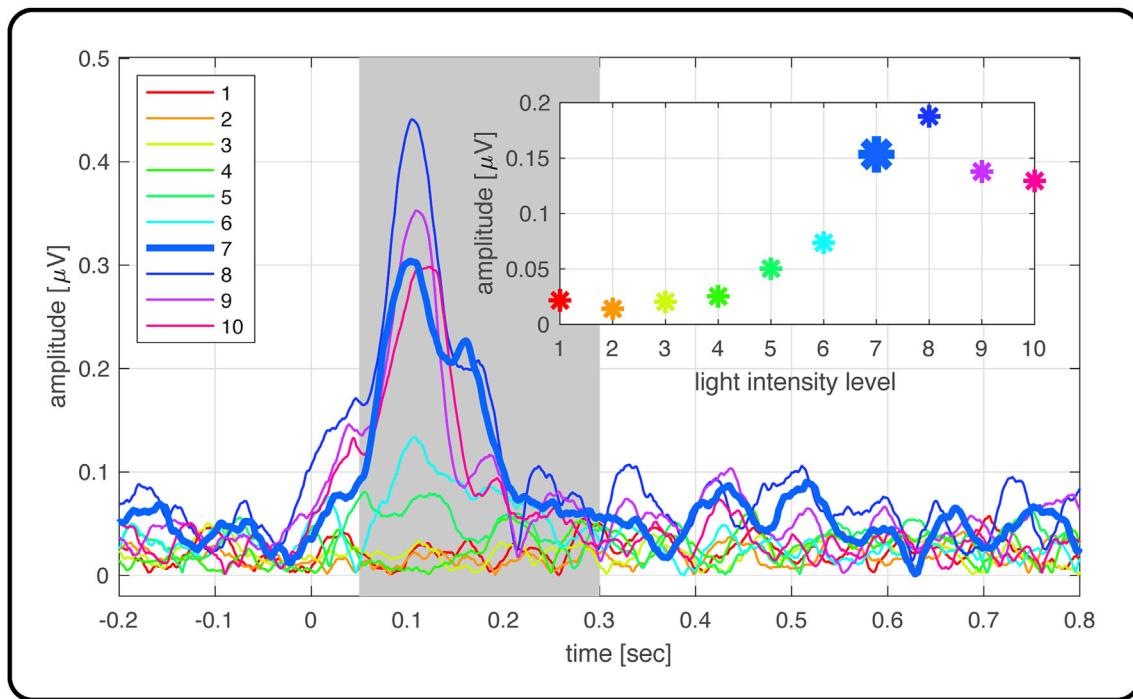


Fig. 4. The procedure to determine the I_{LED} used in ER-NF in one experiment (Sub01b). With the light intensities varying from Level 1 to 10, the VEP waveforms (the mean absolute values of the filtered signal after baseline correction and Hilbert transform) are illustrated with different color and their mean amplitudes in the interval of 0.05–0.3 s are displayed in the upper right plot. In this case, when the light intensity was at Level 7 (as marked by a large blue asterisk), the mean amplitude of VEP was substantially increased, as compared with that at Level 1–6. So, a duty cycle of 0.128 (Level 7) was selected as the I_{LED} used for ER-NF in this experiment.

6 ER-NF sessions. A modulation function to describe how the alpha power at the P_{rest} varies with respect to the phase index (time lag) Φ was estimated to illustrate ER-NF's modulation effect at different alpha phases. A 1024-point Fast Fourier transform (FFT) was applied to estimate the periodicity (i.e., the main frequency) of the modulation function, which was abbreviated as F_{mod} below. If the proposed ER-NF system does really work, the modulation function should be periodical and the period should be same with the alpha wave.

To further illustrate the modulation effect of ER-NF along time, Short-time Fourier Transform (with parameters: a window length of 2 s, a moving step of 20 ms) was applied on modulated EEG during ER-NF. For clarity, we only showed the modulation functions with phase indices Φ corresponding to the maximal and minimal power values in the modulation function (denoted as Φ -Max and Φ -Min). The time-varying spectral power of modulated EEG were averaged over 6 sessions. The peak alpha frequency at each phase index Φ was detected from the time-varying spectral power to depict how ER-NF modulated a subject's peak alpha frequency.

Based on the simple pendulum model, visual stimuli at different phases could either increase or decrease the alpha power and the frequency. Hence, the modulation functions of the alpha power were compared with the P_{rest} (the peak power of resting-state EEG). The modulation function was also compared with the peak power of SSVEP, because ER-NF and SSVEP used visual stimuli to regulate alpha wave in different ways (see Discussion for details).

2.4.2. Group analyses of modulation effects

At the group level, point-wise t -test was firstly applied to detect significant modulation effects (with respect to zero) along time from -20s to 20s. For each experiment, we calculated the instantaneous EEG power with the phase indices Φ -Max and Φ -Min corresponding to the maximal and minimal values in the modulation function. Considering the multiple comparison problem, the cluster-based permutation test (Maris and Oostenveld, 2007) was used with the cluster-threshold setting to 0.05 and permutation for 20000 times.

Then, the correlation between F_{mod} (the main frequency of the periodic modulation functions) and P_{rest} across all subjects was calculated for each experiment to examine whether the periodicity of the modulation function agreed well with the alpha peak. Next, since each subject took participant in the experiment for two times, we calculated the correlation coefficient separately for each experiment. To validate the repeatability of modulation effect, the Fisher r-to-z transformation was used to check whether there was significant difference between correlation coefficients of two experiments.

2.4.3. Comparison of modulation effects among different regulation protocols

To check whether the ER-NF system could significantly increase or decrease the alpha power, we compared the maximum and minimum value in the modulation function with P_{rest} . Further, to compare modulation effects between closed-loop and open-loop visual stimulation protocols, the maximum value of ER-NF's modulation function was also compared with that of SSVEP. Hence, in total four types of EEG alpha power were compared, and they are from the following four protocols:

- ER-NF (Φ -Max): ER-NF with the stimuli delivered at the phase Φ -Max,
- ER-NF (Φ -Min): ER-NF with the stimuli delivered at the phase Φ -Min,
- REST: Resting-state with eyes-open,
- SSVEP: SSVEP in which stimuli were delivered at 10 Hz.

Alpha powers from these four paradigms were compared using repeated measured one-way ANOVA, and then paired-sample t -test is performed between each pair of paradigms for post-hoc tests using Bonferroni adjusted significance level of ($0.05/6 = 8.33 \times 10^{-3}$) per test. Because SSVEP sessions were only available in the second experiments, only the results in the second experiments were compared when SSVEP was in the comparison. More precisely, 20 samples of REST, ER-NF(Φ -Max) or ER-NF(Φ -Min) obtained from the second experiments and all 20 samples for SSVEP were compared, if a comparison included SSVEP.

3. Results

In section 3.1, we firstly took one experiment of one subject as an example to demonstrate the modulation effects of the proposed ER-NF system. Then, group-level results are illustrated in section 3.2. Last, the comparison of modulation effects between different visual stimulation/regulation protocols is presented in section 3.3.

3.1. Single subject analyses of modulation effects

The second experiment of subject 1 (Sub01b) was selected because the results were clear and representative. For Sub01b, the detected F_{rest} was 11 Hz. Hence the bandwidth setting of the Butterworth was 10–12 Hz. The I_{LED} in ER-NF was Level 7 (as explained in Fig. 4).

Fig. 5a shows the modulation effects of ER-NF at F_{rest} when the phase index (time lag) Φ is 0 ms (Φ-Min; blue curve) or 50 ms (Φ-Max; red curve). For both two alpha phases, the modulation took effect from 0 s and plateaued for 20 s until the experiment was switched to the next phase.

Actually, the modulation effects at the phase $\Phi = 50$ ms and $\Phi = 0$ ms respectively achieved the maximal and the minimal values of the

modulation function, and these two phases were marked in red and blue circles in Fig. 5b, which shows the modulation function of the alpha power at $F_{rest} = 11$ Hz against the phase index Φ . The modulation function showed a clear periodicity, and its frequency (F_{mod}) was 10.94 Hz, which was very close to the F_{rest} . For comparison, the power of rest EEG ($P_{rest} = -10.36$ dB) is marked in yellow dash line, and the power of SSVEP (7.76 dB) is marked in violet dots line. Alpha power values of both SSVEP and REST were lower than the maximal power of 9.32 dB in the modulation function of ER-NF, while alpha power values of REST was larger than the minimal power of -20.99 dB in the modulation function.

Further, Fig. 5c shows that, with the increase of the phase index Φ , the peak frequency of the modulated alpha wave moved towards the lower frequency band (marked as white dots in Fig. 5c). The result indicated the proposed ER-NF protocol could not only modulate the power of the alpha wave, but also modulated the peak alpha frequency. In another word, ER-NF could modulate the pattern of alpha wave. It also agreed with the model assumption of the simple pendulum, in which a force exerted at a specific phase could influence not only the amplitude but also the frequency of the simple pendulum.

The modulation effects of ER-NF in all experiments are illustrated in Fig. 6. Similar with Fig. 5b, in each experiment the power of the

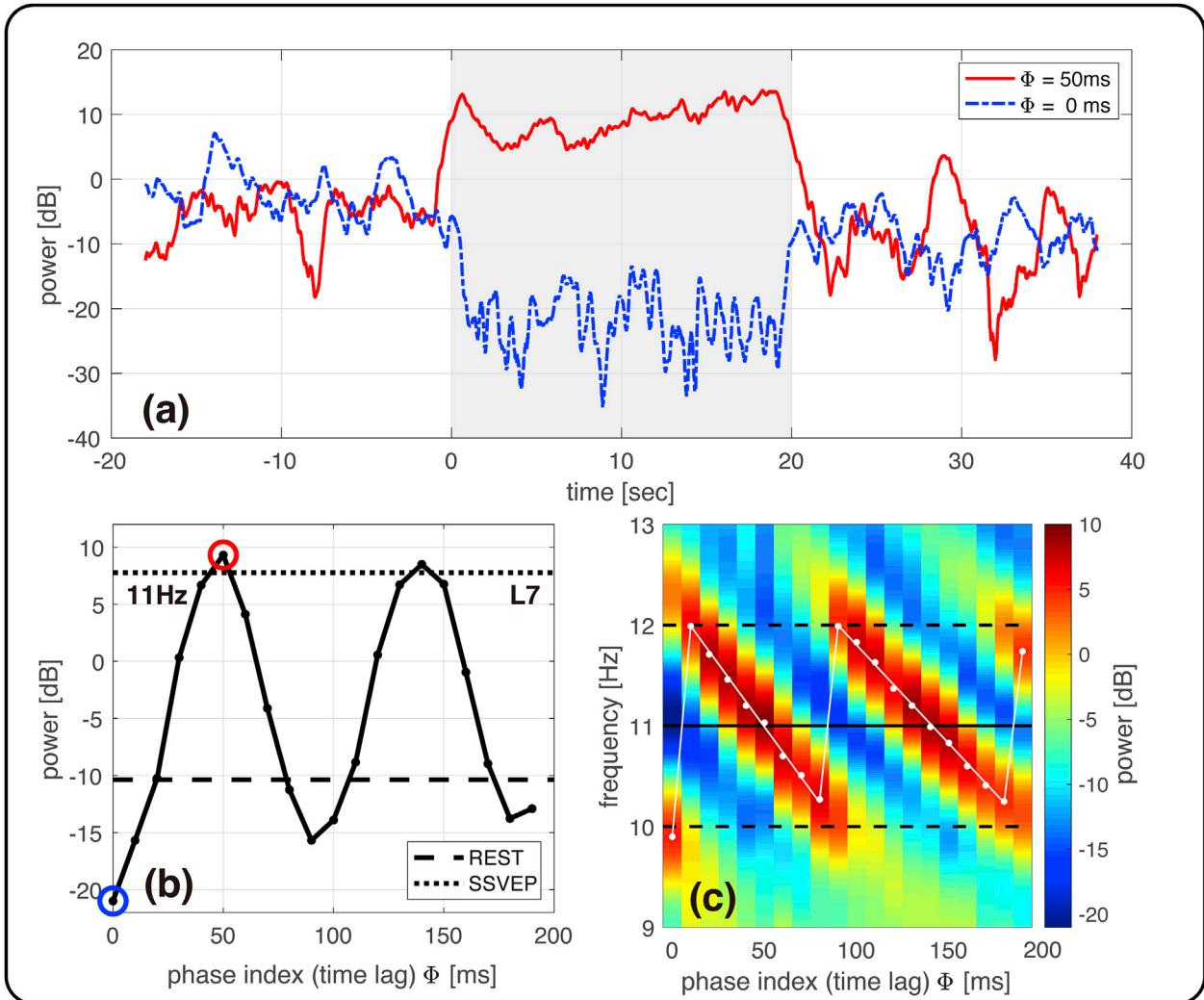


Fig. 5. The modulation effect of ER-NF in the second experiment of Subject 1 (Sub01b). (a) The time response of the modulation effect with the phase index (time lag) Φ corresponding to the maximal ($\Phi = 0$ ms) and minimal ($\Phi = 50$ ms) power in the modulation function. The shadowed interval (0–20 s) indicates the modulation interval with the target phase index Φ . (b) The modulation function against the phase index (time lag) Φ , in which 11 Hz is the value of F_{rest} and L7, short for Level 7, is the value of I_{LED} . The dash line indicates the P_{rest} . The dots line indicates the power of the SSVEP at 10 Hz. The valley and peak of the modulation function are marked by blue and red circles. (c) Power spectra of the alpha waves regulated by visual stimuli delivered at different values of Φ . The solid line and the dash line indicate the central frequency and bandwidth setting of the online filter. The white dots and lines stand for the peak frequencies at different Φ .

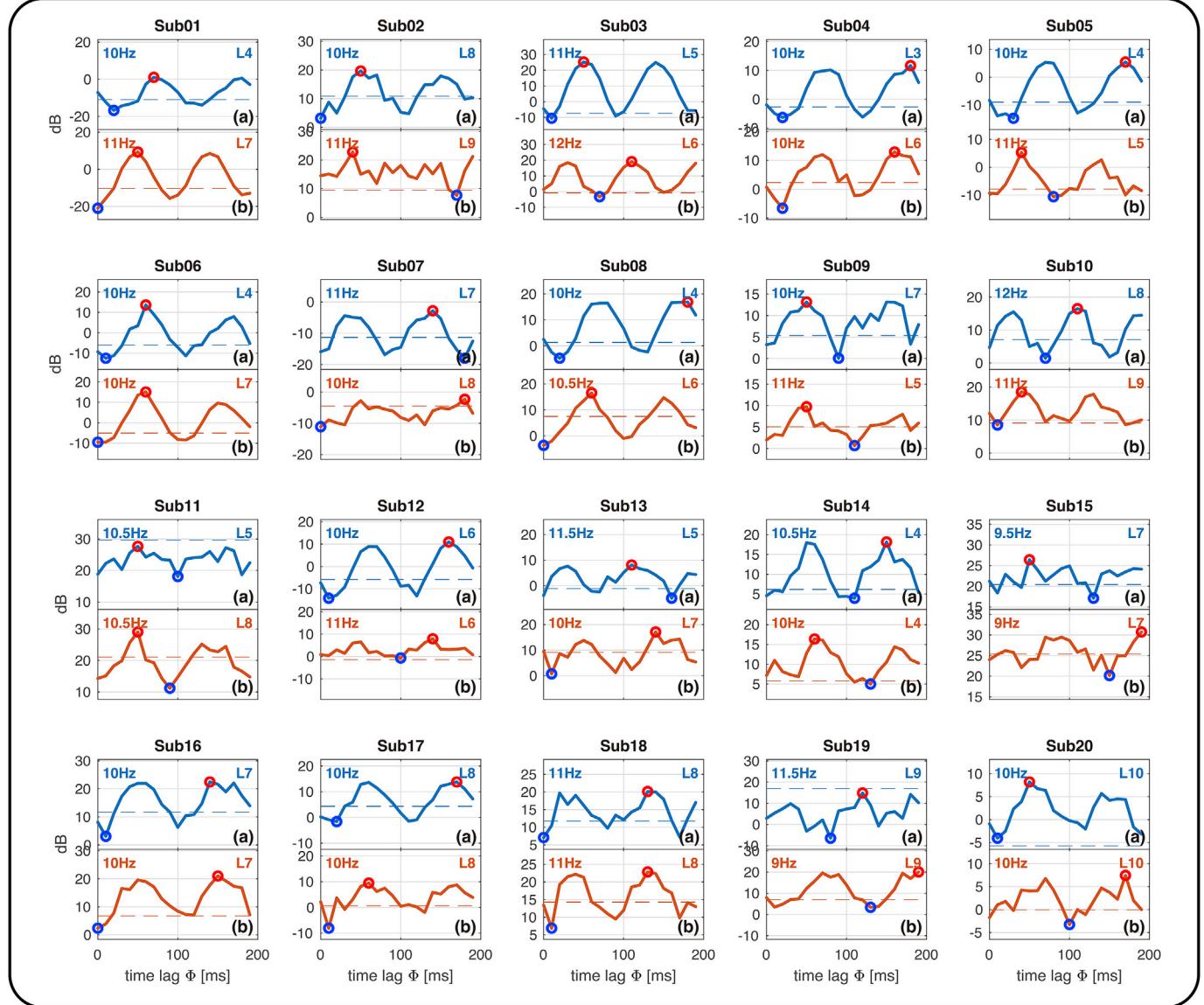


Fig. 6. The modulation effects of the ER-NF in all experiments. For each subject, the power of the alpha wave at its F_{rest} is shown as a power modulation function against the phase index (time lag) Φ , which corresponds to different phases of the alpha wave. The valleys and peaks of the modulation functions are marked by blue and red circles. Two important parameters, F_{rest} and I_{LED} , are labeled in the upper part of each subfigure. The dash line stands for P_{rest} . All subjects took participant in the experiment for two times (blue for the first time, orange for the second time) to test the within-subject repeatability of the ER-NF.

modulated EEG at F_{rest} is shown as a power modulation functions against the phase index Φ . The results for two experiments are respectively shown as blue and orange curves and their parameter settings (F_{rest} and I_{LED}) are also labeled. On average, F_{rest} was 10.35 ± 0.76 Hz, and I_{LED} was Level 6.51 ± 1.75 . It should be stated that, even with the I_{LED} of Level 10, some experiments (Sub02b, Sub07a, Sub07b, Sub09b, Sub10b, Sub18a, Sub20a and Sub20b) did not have clear VEP responses. In these cases, the I_{LED} that evoked the maximal VEP was used in the evaluation part of the experiment.

Similar with the modulation effect of ER-NF of Sub01b in Figs. 5 and 6 shows that clear periodic modulation functions (modulated power of EEG signal at F_{rest} with respect to Φ) could be observed from almost all experiments. The time response of the modulation effect and the power spectra of the alpha waves regulated by visual stimuli delivered at different values of Φ in all experiments is arranged in Figs. S3 and S4. Bad modulation effect, such as that of Sub02b, could be explained by the subject's mental state. The experimenter reported that Sub02 was drowsy in the second experiment (Sub02b), but not in the first experiment

(Sub02a), suggesting that drowsiness may have an impact on the modulation effects of ER-NF. Pairwise comparisons between ER-NF at Φ -Max or Φ -Min and REST can be found in Fig. S5 in the Supplementary Materials.

3.2. Group analyses of modulation effects

The mean modulation effects along the time axis are displayed in Fig. 7a. Using point-wise t -test with cluster-based permutation test to correct the family-wise error rate, it was found that the alpha power of ER-NF (Φ -Max) and ER-NF (Φ -Max) at their F_{rest} were significant different in the whole modulation period. The significant interval, from -0.46 – 20.44 s, was a little bit wider than the modulation window ($[0, 20]$ second, marked in a gray shadow), which was caused by the moving window (2 s) used in power spectral estimation.

F_{mod} (the main frequency of the periodic modulation functions) was significantly correlated with F_{rest} across subjects (black curve in Fig. 7b with $r = 0.86$, $p = 1.99 \times 10^{-12}$ in both experiments 1 and 2. Sub02b

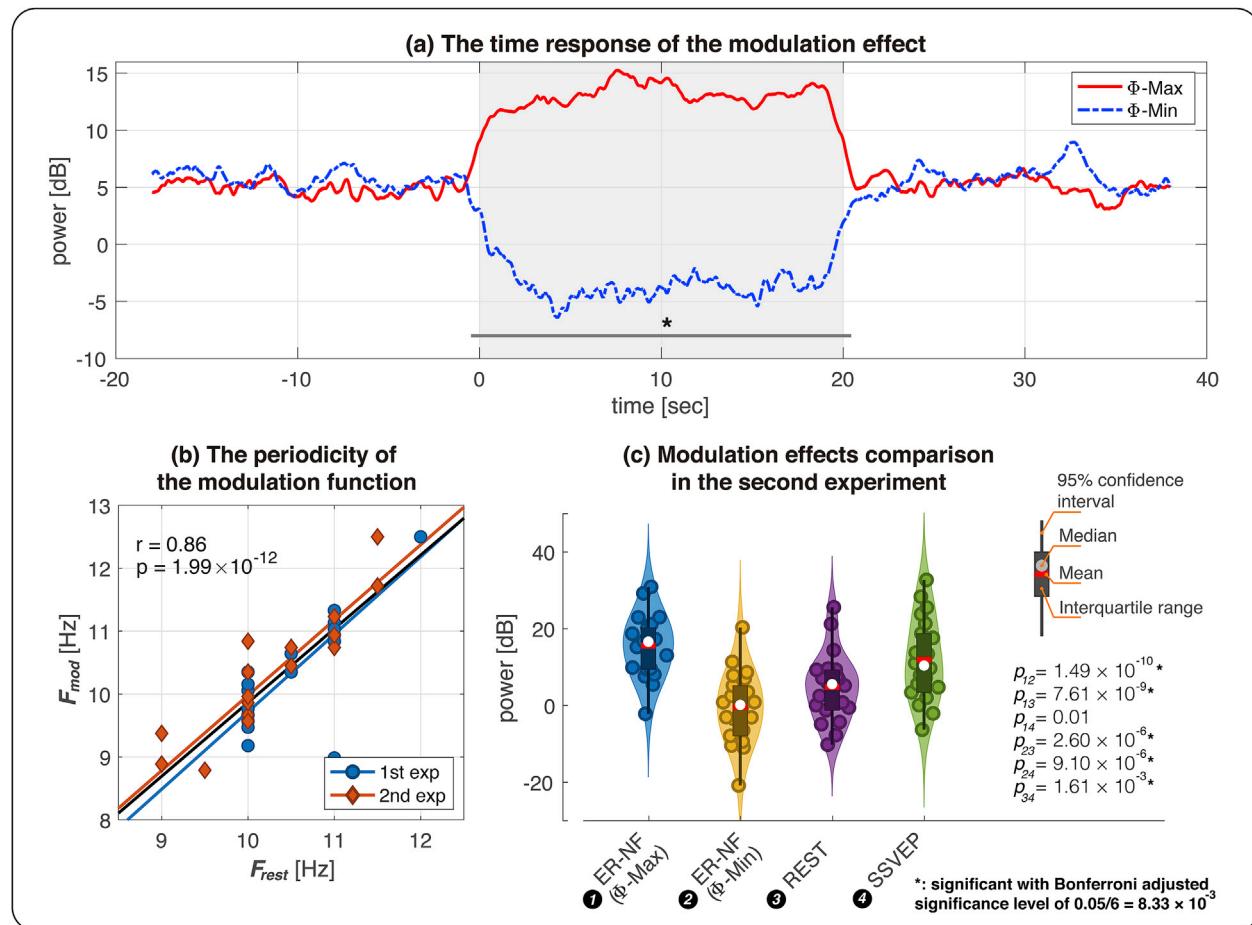


Fig. 7. (a) Time responses of the modulation effects at the phase indices Φ -Max and Φ -Min (corresponding to the maximal and minimal power in the modulation function). The shadowed interval (0–20 s) indicated the modulation interval with the target phase Φ . There is significant difference between the two conditions in the interval of the asterisk from −0.46–20.44 s. (b) The peak frequency of the modulation function is significantly correlated with F_{rest} ($r = 0.86$ and $p = 1.99 \times 10^{-12}$). (c) The violin plot for the alpha powers of ER-NF (Φ -Max), ER-NF (Φ -Min), REST and SSVEP in the second experiment. Since SSVEP data is only recorded in the second experiment, all the statistical analysis is done based on the second experiment. There is significant difference between alpha powers of the four paradigms ($F(3,19) = 13.95$ and $p = 2.11 \times 10^{-15}$). Post-hoc comparisons show that all the pair-wise comparisons, except the comparison between ER-NF (Φ -Max) and SSVEP, had significant difference (i.e., they survived the Bonferroni adjusted significance level of $0.05/6 = 8.33 \times 10^{-3}$).

with $F_{mod} = 19.43$ was excluded in the correlation analysis, because it was detected as an outlier by the criterion that an outlier has a value more than three scaled median absolute deviations away from the median (Leys et al., 2013; see Fig. S6). We also performed correlation analysis separately for the two experiments. The blue and red curves in Fig. 7b show the significant correlation between F_{mod} and F_{rest} in the first experiment ($r = 0.84$, $p = 4.07 \times 10^{-6}$) and the second experiment ($r = 0.91$, $p = 9.39 \times 10^{-8}$), respectively. By comparing the two correlation coefficients with the Fisher r-to-z transformation, we got the z-value $z = -0.84$ ($p = 0.40$), which indicated there was not significant difference between the two experiments.

3.3. Comparison of modulation effects between different stimulation protocols

The power of ER-NFs with phases Φ -Max and Φ -Min were compared with the maximum EEG power in REST and SSVEP (Fig. 7c). Since SSVEP was only recorded in the second experiment, all the statistical comparisons including SSVEP were based on the second experiment. On average, the power for ER-NF (Φ -Max), ER-NF (Φ -Min), REST and SSVEP was 15.46 dB, −0.55 dB, 4.68 dB and 11.78 dB. There is a significant difference among the four protocols ($F(3,19) = 13.95$ and $p = 2.11 \times 10^{-15}$). Paired-sample t-test results are provided in Fig. 7c. Results showed that, all the pair-wise comparisons, except the comparison between ER-NF

(Φ -Max) and SSVEP, had significant difference, which shows the capability of the proposed ER-NF protocol in the regulation of alpha wave. Note that, no sample was excluded as outlier in this comparison.

When comparing modulation effects, REST stands for the resting state EEG with eyes-open. The paired-sample t-test result between ER-NF (Φ -Max) and the resting state EEG with eyes-closed was not significant ($p = 0.09$), so the result of EEG with eyes-closed was not illustrated in Fig. 7c to keep the result concise and clear. As compared with REST, the proposed ER-NF either increased or decreased the alpha power at the frequency F_{rest} ($p = 7.61 \times 10^{-9}$ for ER-NF (Φ -Max) vs. REST and $p = 2.60 \times 10^{-6}$ for ER-NF (Φ -Min) vs. REST and), depending on the stimulation phase (i.e., alpha phase at which the visual stimuli were delivered). In the experiments shown in Fig. 6, we got ER-NF (Φ -Max) > REST > ER-NF (Φ -Min) in the alpha power in most cases. But there were some exceptions: the whole modulation functions for Sub11a and Sub19a were below P_{rest} , and the whole modulation functions for Sub12b and Sub20a were above P_{rest} . In Fig. S5 of the Supplementary Materials, we provided more detailed information about the comparison between REST and ER-NF (Φ -Max/ Φ -Min). SSVEP also evoked higher EEG power than P_{rest} ($p = 1.61 \times 10^{-3}$). With the same I_{LED} , ER-NF (Φ -Max) could evoke marginally significantly higher alpha power than SSVEP ($p = 0.01$, which cannot survive the Bonferroni correction for multiple comparison in the post-hoc tests of ANOVA), while ER-NF (Φ -Min) could achieve significantly lower alpha power than

SSVEP ($p = 9.10 \times 10^{-6}$).

4. Discussion and conclusion

A proof-of-concept system of a new Externally-Regulated Neurofeedback (ER-NF) protocol was designed, implemented, and validated in this study. The main novelty of the ER-NF system is the fact that it is able to regulate the alpha wave by continuously delivering visual stimuli at a specific alpha phase. By using this new ER-NF protocol, users do not need to learn how to regulate their brain activities and the modulation effects are more consistent and flexible.

4.1. Externally-regulated vs. self-regulated neurofeedback

The efficacy of traditional neurofeedback protocols depends on the users' capability to regulate their brain activities in a self-paced voluntary manner. Because of the substantial inter-individual variability in user's skills and experience in self-regulation, the effects of neurofeedback vary hugely. To ensure all users could benefit from neurofeedback, a new EEG regulation method, which is free of user's active participation and training, is desired.

The proposed new ER-NF protocol uses phase-guide visual stimulation to regulate EEG, and it does not rely on users' active participation. As a result, the new protocol can achieve consistently good regulation performance for all users. It is important to note that, external stimuli are not delivered to users according to the operators' predefined specification (such as the delivery time in evoked potential experiments), and they are actually guided by users' EEG but without users' intention. In another word, users do not need to intentionally regulate their brain activities in the new protocol, because the system can decode their brain activities (i.e., to estimate the alpha phase) and accordingly generate external stimuli for users.

An important advantage brought by the ER-NF protocol is that no training is required for new users. In the classical neurofeedback, training is an essential part. Normally, users have to be trained for several weeks or even a couple of months to be able to "voluntarily control" their EEG (Grzelier, 2014). But in the proposed ER-NF system, users need not regulate EEG by themselves, so that training is not necessary. In this study, users were not trained to regulate their EEG, but their alpha waves were effectually regulated even without any training.

The new ER-NF protocol also has advantages in its capability to modulate alpha wave in a flexible and immediate fashion. The proposed ER-NF can either increase or decrease the alpha power, while SSVEP can only increase the alpha power. Also, note that ER-NF actually used exactly the same stimulation intensity, and it is the timing for stimulation (phase index Φ) that makes the difference in the modulation effect. On the other hand, SSVEP increases alpha power by increasing the stimulus intensity, but higher stimulus intensity can cause users' discomfort. In addition, SSVEP evokes EEG responses at the target frequency and its harmonics, while ER-NF affects a user's individualized alpha frequency band. Last, ER-NF can achieve its modulation effect immediately. As soon as the stimuli are delivered, the modulation effect is on; once the stimuli are withdrawn, the modulation effect is gone.

ER-NF also has a merit in its rigorously controlled experiment, which is actually double-blind. A double-blind experiment should eliminate subjective, unrecognized biases carried by the subjects and operators. Most of conventional neurofeedback studies are not double-blind, which makes the efficacy of neurofeedback controversial (for example, some studies argued that neurofeedback efficacy actually comes from placebo effect) (Arnold et al., 2013). Therefore, double-blind placebo-controlled studies should be designed to validate the efficacy of neurofeedback (Heinrich et al., 2007). In ER-NF, since visual stimuli continuously delivered at different phases cannot be discriminated by eyes, both the user and the experimenter did not know which alpha phase was being used to generate visual stimulation and whether the user's alpha power would be increased or decreased during the experiment. Hence, the

ER-NF experiments are double-blind and placebo-controlled, which provided a rigorous validation of the good modulation effect of the ER-NF system.

4.2. Comparisons with other neural modulation techniques

We further compare the ER-NF protocol with other two neural modulation techniques that can regulate or change the brain activities: evoked-potential (EP) paradigms and brain stimulation (such as deep brain stimulation [DBS], transcranial magnetic stimulation [TMS], transcranial direct current stimulation [tDCS], transcranial alternating current stimulation [tACS]). Table 1 summarizes the similarities and differences of four neural modulation techniques.

First, it is clear that, ER-NF, EP, and brain stimulation do not need users' active participation or training, while users of the traditional SR-NF have to be trained to master the skills of self-regulation. So, training is an essential part of conventional neurofeedback systems, and its effect greatly determines the efficacy of neurofeedback.

Second, ER-NF, EP, and brain stimulation techniques have different styles of stimuli and different ways to change the brain activities. Both ER-NF and EP use sensory stimulation (visual stimulation in this study) to change brain signals. But the stimulation parameters used in ER-NF are generated based on decoded brain activities, while the stimulation parameters used in EP are pre-defined by operators. Unlike ER-NF and EP, brain stimulation techniques delivered electrical, magnetic, or acoustic stimulation directly on the brain. Normally, the stimulation parameters used in brain stimulation are defined by operators, though some new brain stimulation techniques can also use decoded brain signals or other physiological signals as feedback to adaptively adjust stimulation parameters (Rosin et al., 2011).

Third, ER-NF and other traditional neurofeedback protocols have a closed-loop, which decodes brain activities to regulate brain activities (via computers and/or other devices or by the subject him/herself). In ER-NF, the decoded phase information is directly used to generate the stimulation sequence. On the other hand, in the traditional neurofeedback, the decoded information is just displayed to users so that they can regulate their brain wave accordingly. EP and brain stimulation techniques modulate the brain in an open-loop fashion: computers or devices deliver pre-defined stimuli to users and there is no feedback.

Fourth, these neural modulation techniques can modulate different types of EEG activities. As a new protocol, ER-NF is found to only modulate the peak alpha frequency and alpha amplitude, which are actually covaried. On the other hand, because the conventional SR-NF has been studied for decades, it has been found to be able to modulate a wide range of EEG activities, from rhythms to brain connectivity. EP paradigms can elicit various types of transient EPs (such as VEP) and steady-state EPs (such as SSVEP), which are reflected as changes in the time-domain morphology or in the spectral characteristics of various frequency bands. Different brain stimulations techniques modulate different types of EEG activities, from time-domain EPs to spectral properties. It must be mentioned that, Table 1 only lists some commonly-used or well-known EEG activities and actually the list is far from exhaustive. Because the brain is a highly dynamic and complex system, external stimulation or self-regulation can change a wide variety of EEG activities. For example, all these neural modulation techniques (including the proposed ER-NF) may be able to alter the functional connectivity and network properties of the brain.

Last but not the least, the underlying neuroscience theories for these four neural modulation techniques are different. The traditional SR-NF is based on learning mechanism of operant conditioning (Rosenfeld et al., 1995; Vernon et al., 2003), which suggests that users can be trained to gain some voluntarily control on their brain wave. ER-NF and EP are based on the theory of neurodynamics, which implies that stimuli delivered at different phases would evoke different dynamic behaviors of neural activities (McSayers and Beagley, 1974; Trimble and Potts, 1975; Jervis et al., 1983). Brain stimulation is normally based on the theory of

Table 1

Comparisons between the ER-NF protocol and other neural modulation techniques.

	Proposed Externally-Regulated Neurofeedback	Traditional Self-Regulated Neurofeedback	Evoked-potential paradigms (VEP, SSVEP et al.)	Brain stimulation (DBS/TMS/tDCS/ TACS) ^a
Need users' active participation and training?	No	Yes	No	No
Need external stimulation?	Yes sensory stimulation generated based on decoded EEG	No visual and/or auditory feedback is just used to show brain state information	Yes sensory stimulation predefined by operators	Yes electric/magnetic stimulation predefined by operators and delivered on the brain
Need to decode brain signals to construct a closed-loop?	Yes (closed-loop) brain → computer/device → sensory system → brain	Yes (closed-loop) brain → computer/device → sensory system → brain	No (open-loop) computer/device → sensory system → brain	No (open-loop) computer/device → brain
Modulated EEG activities	peak frequency and power/amplitude (which are covaried) in the alpha band (based on this study)	peak frequency, power/amplitude spatial distribution (such as asymmetry), connectivity, and other features in various frequency bands (Gruzelier, 2014; Yamashita et al., 2017; Baehr et al., 2001; etc.)	Evoke certain response in time domain and time-frequency domain for VEP; magnitude and phase on certain frequency point with its harmonic frequency point for SSVEP (Luck, 2014)	cortical oscillations in various frequency bands/amplitude and latency for ERP components (Kibleur et al., 2017; Fröhlich, 2015; Vossen et al., 2015; Lenoir et al. 2017; etc.)
Neuroscience theory	Neurodynamics (McSayers and Beagley, 1974; Trimble and Potts, 1975; Jervis et al., 1983)	Operant conditioning (Vernon et al., 2003)	Neurodynamics (McSayers and Beagley, 1974; Trimble and Potts, 1975; Jervis et al., 1983)	Neuroplasticity (Huang et al., 2005)

^a A few new brain stimulation techniques have used brain activities or other physiological recordings as feedback signals to construct a closed-loop, such as closed-loop DBS (Rosin et al., 2011). But, in general, most of existing brain stimulation are open-loop and do not record and decode brain activities for feedback.

neuroplasticity. For example, repetitive transcranial magnetic stimulation will induce plasticity of both excitatory and inhibitory synapses (Lenz et al., 2016; Huang et al., 2005; Hamada et al., 2012), and paired associative stimulation is based on the theory of Hebbian learning (a synapse between two neurons is strengthened when they have highly correlated outputs) (Ridding et al., 2003; Stefan et al., 2000).

Although the ER-NF protocol is similar to some other neural modulation techniques in some aspects, it also has uniqueness and key improvements. For example, the ER-NF protocol is similar to some new closed-loop brain stimulation systems, such as EEG-guided DBS, which can be used on the patients with essential tremor with movement intention (Herron et al., 2015). The advantage of this ER-NF protocol over closed-loop brain stimulation is that, ER-NF uses natural sensory input and thereby is non-invasive, more accessible and user-acceptable.

4.3. Limitations and future work

The new proof-of-concept system still has some limitations to be overcome.

4.3.1. Modulation effects of ER-NF on alpha wave

The proposed ER-NF protocol is able to modulate the power distribution pattern of alpha wave. The present study mainly analyzed the modulation effects of ER-NF on alpha power at individually-defined alpha frequency point F_{rest} , not in the whole alpha band. It is because the modulation effects are most significant at F_{rest} . Actually, with the increase of the range of alpha wave, the modulated alpha power exhibited a similar but weaker modulation effect. This phenomenon may be caused by the band-limited feedback signal we used. Because we used the phase of a band-limited EEG signal (which was bandpass-filtered with cutoff frequencies of $F_{rest} \pm 1$ Hz) as feedback to modulate the alpha wave, the modulate effect may be only evident in the frequency range of $F_{rest} \pm 1$ Hz. In Fig. S1 and Fig. S2 of the Supplementary Materials, we showed the modulation effects of ER-NF in the whole alpha band, and then tried to explain why ER-NF is more effective at individually-defined alpha frequency points. Further, it can be seen from Fig. 5 that ER-NF can modulate the power distribution pattern of alpha wave. When we evaluated the modulation effect at one specific frequency (F_{rest}), the modulation effect was exhibited to be exerted on the alpha power. But we can see from the phase depending spectral power in Fig. 5c that the proposed

ER-NF modulates peak alpha frequency as well. With the increase of the phase index Φ , the peak frequency of the modulated EEG moved towards the lower frequency band (marked as white dots in Fig. 5c for Sub01b and Fig. S4 for all the experiments in the supplementary materials). These results suggested that the modulation effects of the proposed ER-NF protocol on alpha wave could be complicated because alpha power and frequency are modulated interdependently. Considering the fact that the modulation effect on alpha power is weaker if alpha power is calculated in a wider frequency large, it is also possible that ER-NF mainly modulates the peak alpha frequency. This speculation should be examined in future by more specifically designed experiments (for example, to check how different bandwidths of bandpass filtering influence the modulation effect) and more sophisticated mathematical models.

4.3.2. Mathematical models

In Fig. 1, we used a simple model of pendulum to illustrate the principle of ER-NF. Such an analogy is easy to understand but may oversimplify the dynamic behaviors of alpha wave. For example, most neurodynamic models describe the alpha oscillation as a limit cycle attractor (Huang et al., 2011; Freeman, 2015; Acedo and Morano, 2013), not the simple pendulum motion without damping. In future, we should use mathematical functions to describe how alpha power and frequency are modulated by external visual stimuli delivered at a specific phase. Some sophisticated models, such as neural mass models (Jansen and Rit, 1995) and time-series models (Vijayan et al., 2015), could be used to provide a more accurate description and a comprehensive understanding of the alpha wave regulated by phase-guided visual stimuli in ER-NF.

4.3.3. System delay

The proposed system has an inevitable delay stemming from both hardware and software. The hardware delay is caused by data transmission via the EEG amplifier (BrainAmp), which is a constant (60 ms) in our testing. The software delay comes from the phase delay of the online causal Butterworth bandpass filter used to separate alpha wave. The Butterworth filter used is an Infinite Impulse Response (IIR) filter, which introduced different phase delay values at different frequencies. If the F_{rest} is 10 Hz, the maximum delay for the 2-order Butterworth bandpass filter (9–11 Hz) is 247 ms. The filter parameters (such as the filter order and the bandwidth) determine the tradeoff between quality phase estimation of the alpha wave and the filter delay. A higher filter order would

increase the delay, and a wider bandwidth would decrease the accuracy for the phase estimation. Hence, the system delay from both the software and hardware added up to more than 300 ms, which is more than 3 alpha cycles. That is, the alpha phase we estimate is actually the alpha phase at 3 cycles before. Due to the unpredictable dynamic behavior of the alpha wave, the delay of the system may have an influence on the modulation effect. If we can estimate the alpha phase with shorter system delay, the modulation effect may be further improved.

4.3.4. Extensions to other EEG rhythms and other sensory stimuli

This work manipulated alpha wave by visual stimulation. In principle, this ER-NF system can be extended by modulating other EEG rhythms (beta, theta, etc.) and/or using other types of sensory stimulation (such as auditory and somatosensory inputs). We selected alpha band and visual stimulation in this study because alpha wave is strong in EEG and visual stimulation is easy to generate and easy to evoke EEG changes. If we extend the ER-NF system to other EEG rhythms and sensory stimulation, we need overcome the difficulties in precisely and promptly detecting phases of EEG rhythms and in generating user-dependent and phase-guided stimulation sequence. Further, the real time alpha phase estimation technique can also be potentially used in the modulation of sensory perception.

4.3.5. Modulation effects on cognitive and behavioral states

We have used converging results to demonstrate the efficacy of ER-NF in regulating the power and frequency of alpha wave. However, a relationship between alpha wave and mental states does not mean mental states can be altered by modulating alpha wave. Because we did not collect any behavioral and cognitive variables, it is still unknown whether the new system holds the capability to modulate users' behavior and cognition. Alpha wave is closely related to many cognitive and behavioral states, such as attention (Aftanas and Golocheikine, 2001; Klimesch et al., 1998; Klimesch, 2012) and perception (Nunn and Osselton, 1974; Tu et al., 2016; Peng et al., 2015). We believe the proposed ER-NF could regulate perceptual and cognitive variables by modulating alpha oscillation because some other neuromodulation studies have revealed the causal link between alpha wave and perception or cognition. Non-invasive brain stimulation methods, such as TMS, TDCS, and TACS, have shown that, they can modulate the alpha wave and in turn influence perception, cognition and behavior (Herrmann et al. 2013; and Kuo et al. 2012), implying a causal relationship between alpha wave and behavior. Since the proposed ER-NF method can also modulate alpha oscillation, it should be able to change certain domains of cognition or behavior. Because the proposed ER-NF method can provide a more individualized and consistent way to modulate alpha wave than non-invasive brain stimulation, we expect it could achieve better performance in modulating cognition and behavior. Of course, the actual effect of ER-NF on users' behavior and cognition should be rigorously examined by well-designed experiments, large-scale validation, randomized trials, and longitudinal study, and be compared with other types of mainstream and advanced neurofeedback techniques.

In summary, we proposed a new ER-NF protocol, which uses phase-guided external stimulation to regulate brain activities, so that no users' active participation or training is required. The modulation effects of the system can be reliably observed from almost all users in a double-blind test. Therefore, the proposed ER-NF system is an important step towards addressing the neurofeedback inefficiency problem and holds great potential to more reliably and flexibly modulate various domains of cognition and behavior.

Acknowledgements

This work was supported by National Natural Science Foundation of China (No.61701316), Science, Technology and Innovation Commission of Shenzhen Municipality Technology Fund (No. JCYJ20170818093322718), Shenzhen Peacock Plan (No.

KQTD2016053112051497), National Natural Science Foundation of China (No.81871443), and China Postdoctoral Science Foundation Grant (No. 2018M643185). None of the authors have potential conflicts of interest to be disclosed.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.neuroimage.2019.01.072>.

References

- Acedo, L., Moraño, J.A., 2013. Brain oscillations in a random neural network. *Math. Comput. Model.* 57 (7–8), 1768–1772.
- Aftanas, L.I., Golocheikine, S.A., 2001. Human anterior and frontal midline theta and lower alpha reflect emotionally positive state and internalized attention: high-resolution EEG investigation of meditation. *Neurosci. Lett.* 310 (1), 57–60.
- Albrecht, B., Uebel-von Sandersleben, H., Gevensleben, H., Rothenberger, A., 2015. Pathophysiology of ADHD and associated problems-starting points for NF interventions? *Front. Hum. Neurosci.* 9, 359.
- Alkoby, O., Abu-Rmileh, A., Shriki, O., Todder, D., 2018. Can we predict who will respond to neurofeedback? A review of the inefficacy problem and existing predictors for successful EEG neurofeedback learning. *Neuroscience* 378, 155–164.
- Arnold, L.E., Lofthouse, N., Hersch, S., Pan, X., Hurt, E., Bates, B., Kassouf, K., Moone, S., Grantier, C., 2013. EEG neurofeedback for ADHD: double-blind sham-controlled randomized pilot feasibility trial. *J. Atten. Disord.* 17 (5), 410–419.
- Baehr, E., Rosenfeld, J.P., Baehr, R., 2001. Clinical use of an alpha asymmetry neurofeedback protocol in the treatment of mood disorders: follow-up study one to five years post therapy. *J. Neurother.* 4 (4), 11–18.
- Coben, R., Linden, M., Myers, T.E., 2010. Neurofeedback for autistic spectrum disorder: a review of the literature. *Appl. Psychophysiol. Biofeedback* 35 (1), 83.
- Doehnert, M., Brandeis, D., Straub, M., Steinhausen, H.C., Drechsler, R., 2008. Slow cortical potential neurofeedback in attention deficit hyperactivity disorder: is there neurophysiological evidence for specific effects? *J. Neural. Transm.* 115 (10), 1445–1456.
- Freeman, W.J., 2015. Mechanism and significance of global coherence in scalp EEG. *Curr. Opin. Neurobiol.* 31, 199–205.
- Fröhlich, F., 2015. Experiments and models of cortical oscillations as a target for noninvasive brain stimulation. *Prog. Brain Res.* 222, 41–73.
- Fuchs, T., Birbaumer, N., Lutzenberger, W., Gruzelier, J.H., Kaiser, J., 2003. Neurofeedback treatment for attention-deficit/hyperactivity disorder in children: a comparison with methylphenidate. *Appl. Psychophysiol. Biofeedback* 28 (1), 1–12.
- Gruzelier, J.H., 2014. EEG-neurofeedback for optimising performance. III: a review of methodological and theoretical considerations. *Neurosci. Biobehav. Rev.* 44, 159–182.
- Hamada, M., Murase, N., Hasan, A., Balaratnam, M., Rothwell, J.C., 2012. The role of interneuron networks in driving human motor cortical plasticity. *Cerebr. Cortex* 23 (7), 1593–1605.
- Hammond, D.C., 2007. What is neurofeedback? *J. Neurother.* 10 (4), 25–36.
- Heinrich, H., Gevensleben, H., Strehl, U., 2007. Annotation: neurofeedback-train your brain to train behaviour. *JCPP (J. Child Psychol. Psychiatry)* 48 (1), 3–16.
- Herrmann, C.S., Rach, S., Neuling, T., Strüber, D., 2013. Transcranial alternating current stimulation: a review of the underlying mechanisms and modulation of cognitive processes. *Front. Hum. Neurosci.* 7, 279.
- Herron, J., Denison, T., Chizeck, H.J., 2015. April. Closed-loop DBS with movement intention. *Int IEEE EMBS Conf Neural Eng* 844–847.
- Holtmann, M., Sonuga-Barke, E., Cortese, S., Brandeis, D., 2014. Neurofeedback for ADHD: a review of current evidence. *Child Adolesc Psychiatr Clin N Am* 23 (4), 789–806.
- Huang, G., Zhang, D., Meng, J., Zhu, X., 2011. Interactions between two neural populations: a mechanism of chaos and oscillation in neural mass model. *Neurocomputing* 74 (6), 1026–1034.
- Huang, Y.Z., Edwards, M.J., Rounis, E., Bhatia, K.P., Rothwell, J.C., 2005. Theta burst stimulation of the human motor cortex. *Neuron* 45 (2), 201–206.
- Jansen, B.H., Rit, V.G., 1995. Electroencephalogram and visual evoked potential generation in a mathematical model of coupled cortical columns. *Biol. Cybern.* 73 (4), 357–366.
- Jasper, H., Shagass, C., 1941. Conditioning of the occipital alpha rhythm in man. *J. Exp. Child Psychol.* 28 (5), 373.
- Jervis, B.W., Nichols, M.J., Johnson, T.E., Allen, E., Hudson, N.R., 1983. A fundamental investigation of the composition of auditory evoked potentials. *IEEE Trans. Biomed. Eng.* 30 (1), 43–50.
- Kibleur, A., Polosan, M., Favre, P., Rudrauf, D., Bougerol, T., Chabardès, S., David, O., 2017. Stimulation of subgenual cingulate area decreases limbic top-down effect on ventral visual stream: a DBS-EEG pilot study. *Neuroimage* 146, 544–553.
- Klimesch, W., Doppelmayr, M., Russegger, H., Pachinger, T., Schwaiger, J., 1998. Induced alpha band power changes in the human EEG and attention. *Neurosci. Lett.* 244 (2), 73–76.
- Klimesch, W., 2012. Alpha-band oscillations, attention, and controlled access to stored information. *Trends Cognit. Sci.* 16 (12), 606–617.
- Kuo, M.F., Nitsche, M.A., 2012. Effects of transcranial electrical stimulation on cognition. *Clin. EEG Neurosci.* 43 (3), 192–199.

- Leys, C., Ley, C., Klein, O., Bernard, P., Licata, L., 2013. Detecting outliers: do not use standard deviation around the mean, use absolute deviation around the median. *J. Exp. Soc. Psychol.* 49 (4), 764–766.
- Lenz, M., Galanis, C., Müller-Dahlhaus, F., Opitz, A., Wierenga, C.J., Szabó, G., Ziemann, U., Deller, T., Funke, K., Vlachos, A., 2016. Repetitive magnetic stimulation induces plasticity of inhibitory synapses. *Nat. Commun.* 7, 10020.
- Linden, D.E., Habes, I., Johnston, S.J., Linden, S., Tatineni, R., Subramanian, L., Sorger, B., Healy, D., Goebel, R., 2012. Real-time self-regulation of emotion networks in patients with depression. *PLoS One* 7 (6), e38115.
- Lofthouse, N., Arnold, L.E., Hersch, S., Hurt, E., DeBeus, R., 2012. A review of neurofeedback treatment for pediatric ADHD. *J. Atten. Disord.* 16 (5), 351–372.
- Lenoir, C., Huang, G., Vandermeeren, Y., Hatem, S.M., Mouraux, A., 2017. Human primary somatosensory cortex is differentially involved in vibrotaction and nociception. *J. Neurophysiol.* 118 (1), 317–330.
- Lubar, J.F., Swartwood, M.O., Swartwood, J.N., O'Donnell, P.H., 1995. Evaluation of the effectiveness of EEG neurofeedback training for ADHD in a clinical setting as measured by changes in TOVA scores, behavioral ratings, and WISC-R performance. *Biofeedback Self. Regul.* 20 (1), 83–99.
- Luck, S.J., 2014. An Introduction to the Event-Related Potential Technique. MIT press.
- Maris, E., Oostenveld, R., 2007. Nonparametric statistical testing of EEG-and MEG-data. *J. Neurosci. Methods* 164 (1), 177–190.
- McSayers, B., Beagley, H.A., 1974. Objective evaluation of auditory evoked EEG responses. *Nature* 251, 608–609.
- Mihara, M., Hattori, N., Hatakenaka, M., Yagura, H., Kawano, T., Hino, T., Miyai, I., 2013. Near-infrared spectroscopy-mediated neurofeedback enhances efficacy of motor imagery-based training in poststroke victims: a pilot study. *Stroke* 44 (4), 1091–1098.
- Monastra, V.J., Monastra, D.M., George, S., 2002. The effects of stimulant therapy, EEG biofeedback, and parenting style on the primary symptoms of attention-deficit/hyperactivity disorder. *Appl. Psychophysiol. Biofeedback* 27 (4), 231–249.
- Moriyama, T.S., Polanczyk, G., Caye, A., Banaschewski, T., Brandeis, D., Rohde, L.A., 2012. Evidence-based information on the clinical use of neurofeedback for ADHD. *Neurotherapeutics* 9 (3), 588–598.
- Nunn, C.M.H., Osselton, J.W., 1974. The influence of the EEG alpha rhythm on the perception of visual stimuli. *Psychophysiology* 11 (3), 294–303.
- Peng, W., Babiloni, C., Mao, Y., Hu, Y., 2015. Subjective pain perception mediated by alpha rhythms. *Biol. Psychol.* 109, 141–150.
- Ridding, M.C., Uy, J., 2003. Changes in motor cortical excitability induced by paired associative stimulation. *Clin. Neurophysiol.* 114 (8), 1437–1444.
- Rosenfeld, J.P., Cha, G., Blau, T., Gotlib, I.H., 1995. Operant (biofeedback) control of left-right frontal alpha power differences: potential neurotherapy for affective disorders. *Biofeedback Self-Regul.* 20, 241–258.
- Rosin, B., Slovik, M., Metelman, R., Rivlin-Etzion, M., Haber, S.N., Israel, Z., Vaadia, E., Bergman, H., 2011. Closed-loop deep brain stimulation is superior in ameliorating parkinsonism. *Neuron* 72 (2), 370–384.
- Stefan, K., Kunesch, E., Cohen, L.G., Benecke, R., Classen, J., 2000. Induction of plasticity in the human motor cortex by paired associative stimulation. *Brain* 123 (3), 572–584.
- Tan, G., Thornby, J., Hammond, D.C., Strehl, U., Canady, B., Arnemann, K., Kaiser, D.A., 2009. Meta-analysis of EEG biofeedback in treating epilepsy. *Clin. EEG Neurosci.* 40 (3), 173–179.
- Trimble, J.L., Potts, A.M., 1975. Ongoing occipital rhythms and the VER. I. Stimulation at peaks of the alpha-rhythm. *Invest. Ophthalmol. Vis. Sci.* 14 (7), 537–546.
- Tu, Y., Zhang, Z., Tan, A., Peng, W., Hung, Y.S., Moayedi, M., Iannetti, G.D., Hu, L., 2016. Alpha and gamma oscillation amplitudes synergistically predict the perception of forthcoming nociceptive stimuli. *Hum. Brain Mapp.* 37 (2), 501–514.
- Vernon, D., Egner, T., Cooper, N., Compton, T., Neilands, C., Sheri, A., Gruzelier, J., 2003. The effect of training distinct neurofeedback protocols on aspects of cognitive performance. *Int. J. Psychophysiol.* 47 (1), 75–85.
- Vijayan, A.E., Sen, D., Sudheer, A.P., 2015. EEG-based emotion recognition using statistical measures and auto-regressive modeling. *Int. IEEE CICT Conf* 587–591.
- Vossen, A., Gross, J., Thut, G., 2015. Alpha power increase after transcranial alternating current stimulation at alpha frequency (α -tACS) reflects plastic changes rather than entrainment. *Brain stimulation* 8 (3), 499–508.
- Yamashita, A., Hayasaka, S., Kawato, M., Imamizu, H., 2017. Connectivity neurofeedback training can differentially change functional connectivity and cognitive performance. *Cerebr. Cortex* 27 (10), 4960–4970.



Using a new phase-locked visual feedback protocol to affirm simpler models for alpha dynamics



Xingyi Jin ^{a,b,1}, Zhiguo Zhang ^{a,b,c,d,1}, Li Zhang ^{a,b}, Linling Li ^{a,b}, Gan Huang ^{a,b,*}

^a School of Biomedical Engineering, Health Science Center, Shenzhen University, Shenzhen, Guangdong 518060, China

^b Guangdong Provincial Key Laboratory of Biomedical Measurements and Ultrasound Imaging, Shenzhen, Guangdong 518060, China

^c Marshall Laboratory of Biomedical Engineering, Shenzhen University, Shenzhen 518060, China

^d Peng Cheng Laboratory, Shenzhen, Guangdong 518055, China

ARTICLE INFO

Keywords:

Endogenous alpha oscillations
Exogenous visual stimulus
Phase-locked visual feedback (PLVF)
Neural modulation

ABSTRACT

Alpha band oscillations are the most prominent rhythmic oscillations in EEG, which are related to various types of mental diseases, such as attention deficit hyperactivity disorder, anxiety, and depression. However, the dynamics of alpha oscillations, especially how the endogenous alpha oscillations be entrained by exogenous stimulus, are still unclear. Recently, a newly-developed phase-locked visual feedback (PLVF) protocol has shown effectiveness in modulating alpha rhythm, which provides empirical evidence for the further investigation of the neural mechanism of alpha dynamics. In this work, extensive numerical simulations based on four well-studied models were used to investigate the questions that (1) What kind of dynamic model exhibits a modulation phenomenon of PLVF? (2) What is the dynamic mechanism of PLVF for alpha modulation? (3) Which factors affect the modulation effects in PLVF? The result indicates that the dynamics of endogenous alpha oscillations are close to a simpler dynamic structure, like fixed-point attractor or limit-cycle attractor, which shows a global consistent dynamic behavior at different phases of the alpha oscillation. The further analysis explains the dynamic mechanism of PLVF for amplitude and frequency modulation of the alpha rhythm, as well as the influence of parameter settings in the modulation. All these findings provide a deeper understanding of the endogenous alpha oscillations entrained by exogenous phased locked visual stimulus and lead in turn to the refinement of a control strategy for alpha modulation, which could potentially be used in developing new neural modulation methods for cognitive enhancement and mental diseases treatment.

1. Introduction

Discovered by Hans Berger in 1924, alpha rhythm (8–12 Hz) is the earliest recorded human brain rhythm, which is the most prominent rhythmic oscillation in Electroencephalography (EEG) (Cohen, 2017; Ince et al., 2020). Many studies have shown that alpha rhythm is highly related to several types of cognitive functions (Debener et al., 2006; Hanslmayr et al., 2005), like memory (Hsueh, 2017; Maltseva and Masllobov, 1997) and attentional (Hanslmayr et al., 2011). Abnormal alpha rhythm is often related to various types of mental diseases, such as attention deficit hyperactivity disorder (Butnik, 2005; Fox et al., 2005), anxiety (Hammond, 2005), and depression (Choi et al., 2011). Hence, the modulation of alpha rhythm has raised the fascinating prospect of brain functions enhancement and mental disease treatment. Common

approaches of non-invasive neural modulation, like transcranial electrical or magnetic stimulation (TES or TMS), can directly modulate alpha rhythm by delivering external stimuli. However, the strong electromagnetic artifacts during the exogenous stimulation (Noury and Siegel, 2017; Noury et al., 2016) have made the investigation of the modulation mechanism and the inter-subject response variability difficult, and the large response variability has limited the use of these techniques.

In the previous work, Huang et al. (2019) has proposed phase-locked visual feedback (PLVF) protocol for robust artifact-free alpha rhythm modulation. Stimulated by visual flashing, instead of electrical or magnetic stimulation, no electromagnetic artifact would be induced, which makes it possible for online detecting the phase of alpha rhythm and delivering transient visual stimulation at a specific phase to make a closed-loop alpha rhythm modulation. In result, consistent periodic joint

* Correspondence to: School of Biomedical Engineering, Health Science Center, Shenzhen University, Shenzhen 518060, China.

E-mail address: huanggan1982@gmail.com (G. Huang).

¹ They contributed equally to this paper

amplitude-frequency modulation effects have been shown on all subjects in the resting state with eyes open. More detail about the schema of the PLVF system and modulation results of PLVF in real EEG data would be introduced in Section 2.2 and Fig. 1. However, the real dynamic of alpha oscillation is still somewhat a black box for researchers. The uncertainty of internal dynamic structure and parameters in the real EEG modulation hinders our further exploration of the dynamic mechanism for the alpha modulation. The questions mainly focus on the following three points. Below, the problems have been summarized in three aspects.

Firstly, what kind of dynamic model exhibits a modulation phenomenon of PLVF? Generated by thalamo-cortical interaction (Lopes da Silva et al., 1980; Vijayan and Kopell, 2012; Bollimunta et al., 2011; Schreckenberger et al., 2006), alpha rhythm in both visual and somatosensory cortex propagates from higher-order to lower-order areas (Halgren et al., 2019) and can be detected from the occipital lobe during the resting state and has an increased amplitude when the eyes are closed (Kirschfeld, 2005). Since the discovery of alpha rhythm, various mathematical models have been developed to simulate alpha rhythm. Among these models, Jansen's neural mass model (Jansen et al., 1993; Grimbart and Faugeras, 2006) is widely used for its simplicity in describing the alpha rhythm generation. As a macroscopic-level model for alpha dynamic analysis, the neural mass model exhibits multiple dynamic behaviors with different parameter settings, including fixed-point, limit-cycle, and chaotic strange attractors (Huang et al.,

2011). cccc is generated in neuronal networks as a form of filtered noise. Glass et al. (1993), Palus (1993) and Stam et al. (1999) found that alpha rhythm cannot be distinguished from filtered noise. Others believe alpha rhythm is the product of spontaneous oscillation of the brain network, and that its realization is dominated by regular oscillations accompanied by irregular oscillations (Lopes da Silva et al., 1997). In addition, some evidence showed that EEG signals exhibit chaotic behaviors (Pereira et al., 2005; Zhang, 2017). Hence, the modulation phenomenon of PLVF provides us a probe to investigate the dynamics structs of alpha oscillation.

Secondly, What is the dynamic mechanism of PLVF for alpha modulation? Entrainment of the brain oscillations is commonly used to explain the modulation mechanism of repeated brain stimulation techniques, like rTMS and tACS (Schwab et al., 2006). But the strong electromagnetic artifacts make the investigation difficult. For the proposing of PLVF, the authors assumed the alpha oscillation as a simple pendulum model. In result, the modulation effect on the amplitude of alpha rhythm was the same as expected, but the joint amplitude-frequency modulation was not expected. Furthermore, the undamped simple pendulum model is not stable, and it cannot achieve a stable modulation effect with the PLVF protocol. With the understanding about the dynamics structs of alpha rhythm in the first question, how about the dynamic mechanism to produce the joint amplitude-frequency modulation result, especially how the endogenous alpha oscillations be entrained by exogenous

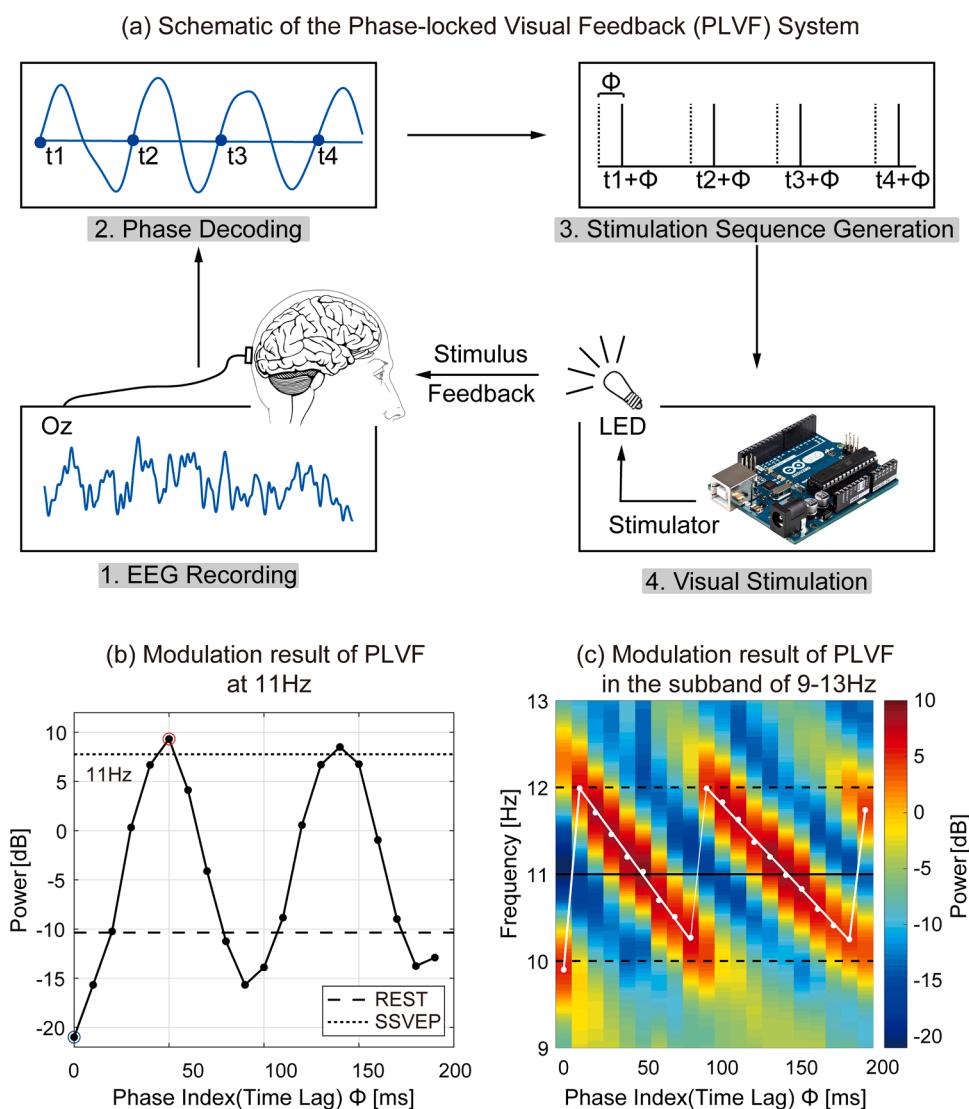


Fig. 1. Schema of PLVF system and modulation results of PLVF in real EEG data: (a) Schema of PLVF system; Raw EEG signals are firstly recorded and filtered, and then the phase of the alpha wave is detected which is used for the generation of visual stimulation sequence (i.e., the exact time to deliver visual stimuli). Finally, visual stimuli were delivered by LED to provide feedback to users. (b) The modulation effect of power shows a sinusoidal-like shape at 11 Hz of the real EEG data, in which the dash line indicates the power of alpha rhythm in resting state with eyes open without modulation, the dots line indicates the power of the SSVEP at 10 Hz. The valley and peak of the modulation function are marked by blue and red circles. (c) Joint amplitude-frequency modulation effect of PLVF in real EEG data (online EEG modulation). As the phase index ϕ varies, the modulation effect at the peak frequency has clear periodicity, in which the peak frequency indicates the frequency with the biggest power in the alpha band.(For interpretation of the references to colour in this figure, the reader is referred to the web version of this article.)

(Reprinted from Huang et al. (2019), Copyright (2019), with permission from Elsevier).

stimulus, is one of the main concerns to investigate the alpha dynamics.

Thirdly, which factors affect the modulation effects in PLVF? Due to the uncertainty of internal dynamic structure and parameters, it is not easy to find the factors which potentially affect the modulation effects. To test the modulation result with different parameters is achievable in the numerical simulation but would be time-consuming in a real EEG experiment. Furthermore, the intra-subject variability caused by the nonstationary internal parameters would make the modulation result noisy. Based on the understanding of the dynamic structure and the modulation mechanism, we can further explore the factors affecting the modulation effect by numerical simulation. The results would in turn lead us to refine the control strategy for alpha modulation.

In this work, focusing on the questions of (1) What kind of dynamic model exhibits a modulation phenomenon of PLVF? (2) What is the dynamic mechanism of PLVF for alpha modulation? (3) Which factors affect the modulation effects in PLVF? Numerical simulations were applied on the neural mass model and three types of typical dynamic attractors including fixed-point, limit-cycle, and chaotic strange attractors to investigate the dynamics mechanism of alpha rhythm modulation by PLVF. The remainder of this paper is structured as follows. Section 2 introduces the PLVF system and the four dynamic models for the simulation of alpha rhythm modulation. Methods to investigate the dynamic mechanism of alpha modulation are described in Section 3. Simulation results are given in Section 4, while discussion and conclusions are provided in Section 5.

2. PLVF system and simulation models

2.1. Closed-loop neural modulation

Common approaches of non-invasive neural modulation, like transcranial electrical or magnetic stimulation (tES or TMS), can directly modulate alpha rhythm by delivering external stimuli. However, strong electromagnetic artifacts during the exogenous stimulation process (Noury and Siegel, 2017; Noury et al., 2016) make real-time monitoring of alpha rhythm difficult, which makes closed-loop modulation difficult. Although many methods were proposed to remove the artifacts online or offline, few of them have been widely accepted (Kasten and Herrmann, 2019). To address this issue, one approach was often adopted that the intermittent stimulation was immediately exerted when intermittent EEG recordings were paused and immediately analyzed (Beliaeva et al., 2021). Take transcranial alternating current stimulation (tACS), several such studies were published. Mansouri et al. (2018) used different phases of pulsed transcranial current to modulate theta and alpha oscillations according to the previous short-time oscillation phase. Zarubin et al. (2020) utilized 1 s tACS after analyzing previous 1 s oscillations to modulate alpha oscillations. Unfortunately, the modulation effects were often hard to interpret, which may be due to many factors such as the inherent delay of the system or the task design. Hence, it is quite hard now for tES or TMS for a real-time stimulation parameters adjustment to modulate the frequency or the phase of the individual neural oscillations to develop a closed-loop system (Frohlich and Townsend, 2021).

2.2. PLVF system

Compared with tES / TMS technique, visual stimuli can modulate the brain rhythm without any electromagnetic artifacts. To achieve closed-loop alpha modulation and realize more precise alpha modulation, Huang et al. (2019) propose a PLVF protocol for alpha rhythm modulation based on a simple pendulum assumption in the previous work. By delivering visual stimuli at different phases, PLVF modulation can induce different amplitude and frequency responses of the alpha rhythm. As shown in Fig. 1a, the schema of the proposed PLVF modulation includes four modules, which are EEG recording, phase decoding, stimulation sequence generation, and visual stimulation. Because the sequence of visual stimuli was generated based on the alpha phase and

then, in turn, modulated the alpha wave, closed-loop control of the alpha wave was formed.

- **Module 1 EEG recording.** The raw EEG signal is online recorded by Self developed C++ programmed from channel Oz with the sampling rate of 5000 Hz and referenced to FCz.
- **Module 2 Phase decoding.** To achieve real-time phase detection, the real-time EEG signal is firstly online filtered by a 2-order alpha-band (8–12 Hz) Butterworth bandpass filter. And then a zero-crossing point detection method is used on the filtered alpha rhythm to identify the positive zero-crossing points with phase $3\pi/2$.
- **Module 3 Stimulation sequence generation.** A certain time lag ϕ (phase index) is introduced to estimate the time point of the alpha rhythm at other phases. Since the alpha rhythm has a period of around 100 ms, it is expected that the modulation effect on the alpha power with a different phase index ϕ also varies in a period of around 100 ms.
- **Module 4. Visual stimulation.** A visual stimulation sequence is generated to be delivered at one fixed phase for each period of the alpha rhythm, which leads to closed-loop alpha rhythm modulation.

In the result shown in Fig. 1b and c, the amplitude of alpha rhythm shows a periodic change with the stimulation phase index ϕ . As phase index ϕ increases, the peak frequency of the modulated alpha rhythm periodically moves toward the lower-frequency bands. A more complete description of the PLVF system and modulation results can be found in Huang et al. (2019).

2.3. Dynamic model of PLVF

The dynamic model of the PLVF system can be described by a differential equation,

$$X' = F(X) + Ku(\tilde{y}, \phi), \quad (1)$$

where X is the state variable as a function of time t , y is the observation representing the recorded EEG signal in the PLVF system, \tilde{y} is the online filtered signal of y , ϕ is the time lag, and K is a vector containing the stimulus direction and intensity. $F(X)$ represents the endogenous evolution of the alpha rhythm, which can be described by different models for simulation. $u(\tilde{y}, \phi)$ is the exogenous stimulus in the PLVF system. As a sum of impulse functions, $u(\tilde{y}, \phi)$ depends on the filtered EEG signal \tilde{y} and the introduced time lag ϕ . The m^{th} positive zero-crossing time point of the filtered signal \tilde{y} is detected as t_m , with $\tilde{y} = 0$ and $\tilde{y}' > 0$. A time lag ϕ is added to t_m to produce the phase at which the stimuli are delivered. Hence $u(\tilde{y}, \phi)$ can be written as

$$u(\tilde{y}, \phi) = \sum_m \delta(t - (t_m + \phi)), \quad (2)$$

where $\delta(t)$ is the unit impulse function with

$$\int_i^j \delta(t) dt = \begin{cases} 1, & i < 0 < j, \\ 0, & \text{otherwise.} \end{cases} \quad (3)$$

2.4. Dynamic models of the alpha rhythm

To further understand the dynamic of the alpha rhythm and the modulation mechanism of the PLVF system, different neural dynamic models are applied in this study. Previous work modeled the alpha rhythm as the motion trajectory of a simple pendulum without damping (Huang et al., 2019). But this is too simple to precisely describe the alpha rhythm, which is unstable with continuously delivered phase-locked stimuli in the PLVF system. To investigate the dynamic structures of alpha rhythm with joint amplitude-frequency modulation (Fig. 1b) in the PLVF system, four types of dynamical models are applied. The first

one is the neural mass model, which has been well studied for the generation of alpha oscillation (Huang et al., 2011). Since oscillation is the basic characteristic of the alpha rhythm, three typical types of oscillatory attractors are also used to simulate the alpha rhythm, which are the linear fixed-point attractor, nonlinear limit-cycle attractor, and chaotic strange attractor.

- Model 1: Neural mass model

At the cellular level, it is well known that the generation of EEG relies on the interactions within or between neurons in the excitatory and inhibitory populations. Jansen's neural mass model (Huang et al., 2011; Jansen and Rit, 1995) can simulate the interaction of excitatory and inhibitory populations. In this model, the neural population is simulated by a population of pyramidal cells, and the output is subject to excitatory and inhibitory synaptic control. The state variable X is a six-dimensional vector, $X = [x_1, x_2, x_3, x_4, x_5, x_6]^T$, with the initial value randomly selected from the interval $x_1 \in [0.1, 0.13]$, $x_2 \in [23.6, 24]$, $x_3 \in [14.5, 18]$, $x_4 \in [-1.4, 1.3]$, $x_5 \in [-14, 10.5]$, $x_6 \in [-108, 110]$, since the attractor located in the region. The output $y = x_2(t) - x_3(t)$, the stimuli intensity $K = [0, 0, 0, 0, 3000, 0]^T$, and

$$F_1(X) = \begin{bmatrix} x_4 \\ x_5 \\ x_6 \\ \text{Sigm}(x_2 - x_3) - 2ax_4 - a^2 x_1 \\ AaC_2 \text{Sigm}(C_1 x_1) - 2ax_5 - a^2 x_2 + Aap \\ BbC_4 \text{Sigm}(C_3 x_1) - 2bx_6 - b^2 x_3 \end{bmatrix}, \quad (4)$$

where $\text{Sigm}(v)$ is the nonlinear sigmoid function,

$$\text{Sigm}(v) = \frac{2e_0}{1 + e^{s(v_0-v)}}, \quad (5)$$

and the parameter settings are

$$\begin{aligned} A &= 3.25mV, \quad a = 100s^{-1} \\ B &= 22mV, \quad b = 50s^{-1} \\ e_0 &= 2.5s^{-1}, s = 0.56mV^{-1}, v_0 = 6mV \\ C_1 &= 1.25C_2 = 4C_3 = 4C_4 = C. \end{aligned}$$

With $p = 200$ and $C = 135$, the output of neural mass model $y = x_2 - x_3$ shows typical alpha-like activity.

- Model 2: Fixed-point attractor

According to the theory that alpha rhythm is generated by the filtering of brain noise, a noise-driven linear point attractor is established. The state variable for this model is a two-dimensional vector, $X = [x_1, x_2]^T$, with the initial value randomly selected from the interval $x_1 \in [-2, 2], x_2 \in [-2, 2]$, since the attractor is located in the region. The output $y = x_1$, the stimulus intensity $K = [3000, 0]^T$, and

$$F_2(X) = \begin{bmatrix} m & n \\ -n & m \end{bmatrix} X + \epsilon, \quad (6)$$

where ϵ is Gaussian white noise with mean 0 and variance 100. The value of m indicates the rate of convergence of the system, and n indicates the rhythm of the oscillation. With $m = -10$ and $n = 70$, the output of the point attractor model $y = x_1$ shows an alpha-like activity with a peak frequency around 11 Hz.

- Model 3: Limit-cycle attractor

According to the theory that alpha rhythm is generated by the spontaneous oscillation of the brain's neural network, a nonlinear limit-cycle attractor with constant angular velocity is established. It

is a simple close orbit with periodic oscillation. The state variable is a two-dimensional vector, $X = [x_1, x_2]^T$, with the initial value randomly selected from the interval $x_1 \in [-1.5, 1.5], x_2 \in [-1.5, 1.5]$, since the attractor is located in the region. The output $y = x_1$, the stimulus intensity $K = [3000, 0]^T$, and

$$F_3(X) = \begin{bmatrix} \frac{kx_1}{r} - kx_1 - cx_2 \\ \frac{kx_2}{r} - kx_2 + cx_1 \end{bmatrix}, \quad (7)$$

where the limit cycle trajectory is a circle with radius $r = \sqrt{x_1^2 + x_2^2}$. The value of k determines the rate of system convergence, and c determines the angular velocity, which is constant. With parameters $k = 10, c = 60$, the output of the limit-cycle attractor $y = x_1$ shows an alpha-like activity.

- Model 4: Chaotic strange attractor

According to the theory that the irregularity of alpha rhythm may be related to the chaotic processes of the brain, the Lorenz attractor is established. This nonlinear chaotic attractor was obtained by meteorologist Edward Lorenz when he studied atmospheric convection in weather forecasting. It is a complex fractal orbit characterized by unstable, driving orbits. The state variable is a three-dimensional vector, $X = [x_1, x_2, x_3]^T$, with the initial value randomly selected from the interval $x_1 \in [-130, 134], x_2 \in [-169, 176], x_3 \in [30, 300]$, since the attractor is located in the region. The output $y = x_3$, the stimulus intensity $K = [0, 0, 100000]^T$, and

$$F_4(X) = \begin{bmatrix} -\sigma x_1 + \sigma x_2 \\ \gamma x_1 - x_2 - x_1 x_3 \\ x_1 x_2 - d x_3 \end{bmatrix}, \quad (8)$$

where $\sigma = 80$ is the Prandtl number, $\gamma = 180$ is the Rayleigh number, and $d = 25$ is a velocity damping constant.

3. Methods

Aiming to understand the alpha dynamics behind the joint amplitude-frequency alpha modulation in the PLVF system, four dynamic models are used for a simulation study. Compared to the chaotic strange attractor and the neural mass model, which dynamic are complex, and the noise-driven fixed-point attractor, which contains noise and makes the explanation of the dynamic difficult, the limit-cycle attractor has a simpler dynamic structure. Hence, it is then used to investigate the underlying dynamic mechanism of PLVF modulation. Finally, we analyze parameters in the model to find out key factors that may affect the results.

3.1. The method of numerical simulation

The Runge-Kutta method is applied in the simulation to solve Eq. (1). The simulation time is 200 s, the sampling rate is 1000 Hz, and the iteration step is 0.001 s. Parameter settings of the four models are shown in Section 2.3. In the first 40 s, there is no stimulation, so that the model can enter a stable state. External stimuli are applied from the 40th second to realize phase-locked modulation. The output signal of the model (simulated EEG signal) is recorded for power spectrum analysis, which is performed within 100–200 s of the output signal using Welch's method, with a window width of 5000, overlapping of 50%, and frequency ranging from 6 Hz to 15 Hz in steps of 0.01 Hz. All simulation and analyses are carried out using MATLAB R2020a software on a Hewlett-Packard ProDesk computer with an Intel i7-6700 CPU at 3.4 GHz. Considering the different initial values, all in the simulation have been repeated 10 times. For each model, the modulation results of each time in two specific phases are shown to illustrate the influence of the set of the initial values (Neural mass model: $\phi = 50, 120$ ms; Fixed-

point attractor: $\phi = 55, 100$ ms; Limit cycle attractor: $\phi = 80, 100$ ms; Lorenz attractor: $\phi = 70, 100$ ms). Meanwhile, the representative phases for each model are chosen, which corresponds to the inhibition and enhancement modulation effects.

3.2. The analysis of alpha modulation effects

For simplicity of analysis, the limit-cycle attractor is used to investigate the underlying neurodynamic mechanism of the PLVF system. The phase portraits of PLVF modulation at time lag $\phi = 10$ ms and $\phi = 80$ ms are illustrated to explain the dynamic mechanism of amplitude and frequency modulation. The PLVF modulation result is described with the increase of phase index ϕ from 0 to 200 ms. The phase portraits with phase index $\phi = 10, 45, 80, 115, 150$, and 185 ms are used to describe the dynamics of neural modulation with the increase of phase index. The power and frequency modulation with the increase of phase index are explained to summarize the dynamic mechanism of PLVF.

3.3. The influence of parameters setting

A signal simulated by a neural mass model, or a limit-cycle attractor cannot precisely simulate real alpha rhythm due to purity. Considering the complex dynamics of the neural mass model and strange attractor, the dynamics of a noise-driven fixed-point attractor are relatively simple. Therefore, the fixed-point attractor is chosen to investigate the factors affecting the modulation results. Four factors are under investigation, and they are stimulus intensity, the bandwidth of the bandpass filter, the center frequency of the bandpass filter, and the accuracy of phase estimation.

In online EEG modulation, the stimulation intensity is an important parameter, which should be large enough to produce an evident modulation effect, but may make participants feel uncomfortable when it is too large. Hence, we consider stimulation intensity as an important factor to investigate the effects of modulation results. Furthermore, because the peak frequency of alpha rhythms differs between individuals (Grandy et al., 2013), the setting of the bandpass filter (center frequency and filter bandwidth) may have a great influence on the modulation results. Therefore, we examine the influence of a bandpass filter with: (1) fixed center frequency and variable filter bandwidth; and (2) fixed filter bandwidth and variable center frequency. In addition, the accuracy of phase estimation is important to PLVF modulation to ensure the modulation effect because, when stimulated at different phases, alpha rhythms have different dynamic responses and modulation effects (Brandt, 1997). Therefore, the accuracy of phase estimation is another important factor to explore. The initial value is randomly selected from the interval $x_1 \in [-2, 2]$, $x_2 \in [-2, 2]$. The parameter settings are as follows.

(1) Stimulation intensity

In the simulation, the stimulation intensity increases from 0 to 5000 with a step of 1000. The amplitude modulation range is estimated to show the modulation effect of stimulation intensity, which is calculated based on the minimum and maximum amplitude of the alpha rhythm during modulation.

(2) Bandwidth of bandpass filter

In the simulation, the bandwidth of the bandpass filter is set to 6–16 Hz, 7–15 Hz, 8–14 Hz, 9–13 Hz, and 10–12 Hz, with the center frequency fixed at 11 Hz. The frequency modulation range is estimated to show the modulation effect of filter bandwidth, which is calculated based on the minimum and maximum peak frequency during modulation.

(3) Center frequency of bandpass filter

In the simulation, the center frequency increases from 8 Hz to 12 Hz, with an interval of 1 Hz. The filter bandwidth is fixed at 2 Hz. The frequency modulation range is estimated to show the modulation effect of the center frequency, which is also

calculated based on the minimum and maximum peak frequency during modulation.

(4) Accuracy of phase estimate

Since the error in phase estimation is inevitable in real EEG modulation by PLVF, which will increase with the phase index, the phase estimate accuracy is considered in this simulation, with the form

$$\tilde{\Phi} = \Phi(1 + e_i R_d), \quad (9)$$

where Φ is the real phase index, R_d is a random number from a zero-mean normal distribution, and e_i is the error index, which increases from 0 to 1, with a step of 0.01. Here, we use $\tilde{\Phi}$ with the error in the phase estimation instead of Φ in Eq. (1). We set the parameter of the error index from 0 to 1, with an interval of 0.01. Due to the error of the phase estimation, the modulation effect of alpha power is attenuated with the phase index varies. Therefore, we use exponential fitting to fit the power modulation effect, and the fitting attenuation coefficient μ is used to describe the attenuation of power as the phase index increases, which is used to assess the modulation effect of phase estimation accuracy. The fitting formula of the attenuation coefficient is

$$y = \xi e^{\mu x}, \quad (10)$$

where μ is the fitted attenuation coefficient, ξ is a positive real number not equal to 1, the argument x is the phase index, and y is the fitted result.

4. Results

4.1. Simulation results in different models

The modulation effects on the amplitude and frequency of the alpha rhythm are shown in Figs. 2 and 3 for the four types of models.

The modulation result for the neural mass model is illustrated in Fig. 2a, which shows similar joint amplitude-frequency alpha rhythm modulation with the change of the stimulation phase. But this still differs from real EEG modulation. First, the peak frequency of the modulated alpha rhythm does not increase linearly with the increase of the phase index ϕ . Furthermore, the peak of the power modulation effect (black dots in Fig. 2a) is at the lower frequency of the frequency modulation range, while it is in the middle frequency of the online EEG frequency modulation range (Fig. 1b). The nonlinearity of the neural mass model causes globally inconsistent dynamic characteristics, so the same force at different positions in the phase space leads to different effects on the system. The modulation results at different phase indexes ($\phi = 50, 120$) have largely different shapes, as illustrated in the middle of Fig. 2a.

Compared to the neural mass model, the dynamic structures of the fixed-point and limit-cycle attractors are much simpler. The fixed-point attractor mode is a linear model with noise, in which all positions share the same dynamic characteristics. The negative real part of the eigenvalues of the matrix $\begin{bmatrix} m & n \\ -n & m \end{bmatrix}$ shows the system would eventually reach the fixed point. However, the existence of noise prevents the convergence of the solution to the fixed points. The periodic oscillation with alpha rhythm is caused by the imaginary part of the eigenvalues in the simulation model. As illustrated in the middle of Fig. 2b, the modulation results of the fixed-point attractor are closer to the real online EEG modulation result in Fig. 1b. With phase-locked stimuli at different phases, the alpha rhythm induces different modulation results, while the phase portraits of modulation results at phase indexes $\phi = 55, 100$ are too noisy to observe the modulation mechanism.

Without the influence of noise, the modulation result from the limit-cycle attractor is clearer and more regular in Fig. 2c. In the proposed limit-cycle attractor, the noise in the fixed-point model is replaced by a constant outward force, which makes the model nonlinear and more

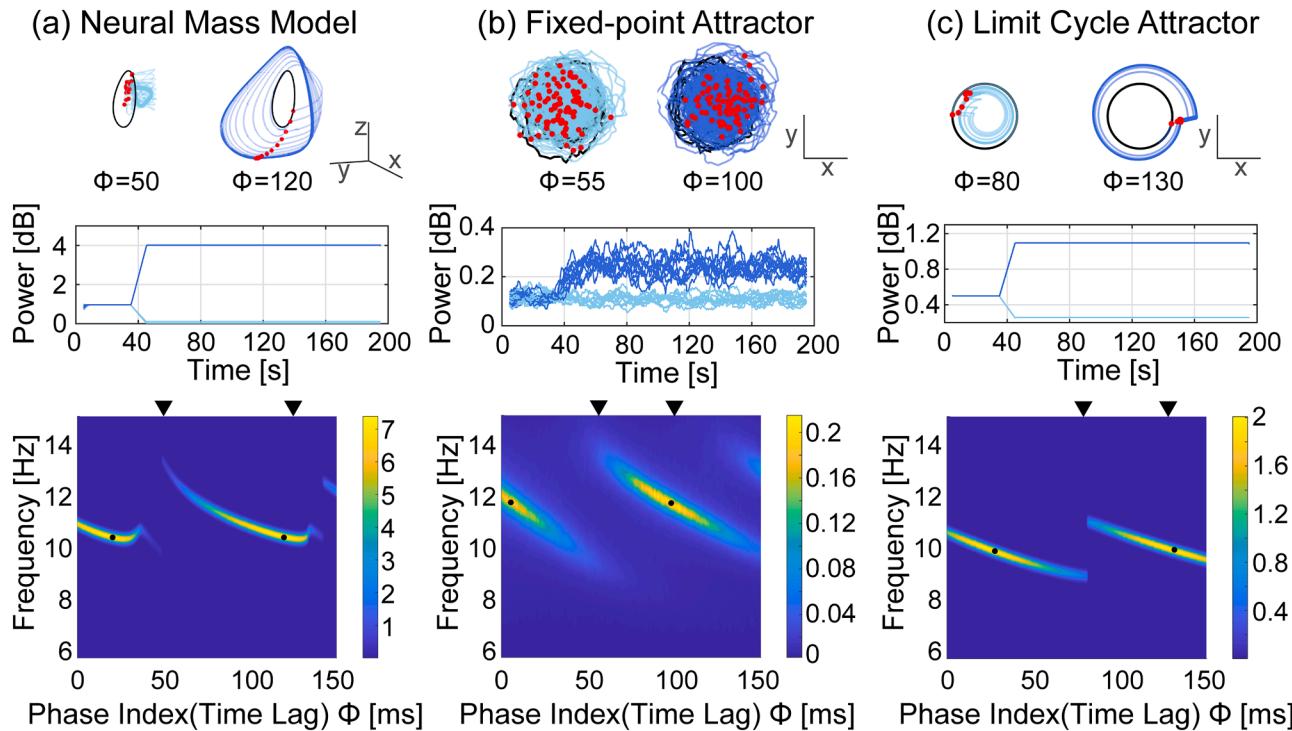


Fig. 2. Modulation results (amplitude and frequency modulation functions against phase index/time lag $\phi = 0 - 150$ ms) of (a) the neural mass model, (b) the fixed-point attractor, and (c) the limit-cycle attractor. Phase portraits at the top raw are modulation trajectories with stimuli at the two different phase indexes (indicated by black triangles at x-coordinates of lower plots). The black curve indicates the resting state of the system from 38 to 40 s before the stimulus. And the blue curve indicates the modulated state of the system from 40 to 48 s after the stimulus. The results in the middle raw show the power changed with time with 10 different initial values and stimuli at the two different phase indexes. Considering a 10 s window is used for smoothing, all the results are shown from 5 to 195 s. Modulation results at the bottom show the joint amplitude and frequency modulation effect and are the average of the results of 10 different initial values. The black dots indicate the peak of the power modulation effect.(For interpretation of the references to colour in this figure, the reader is referred to the web version of this article.)

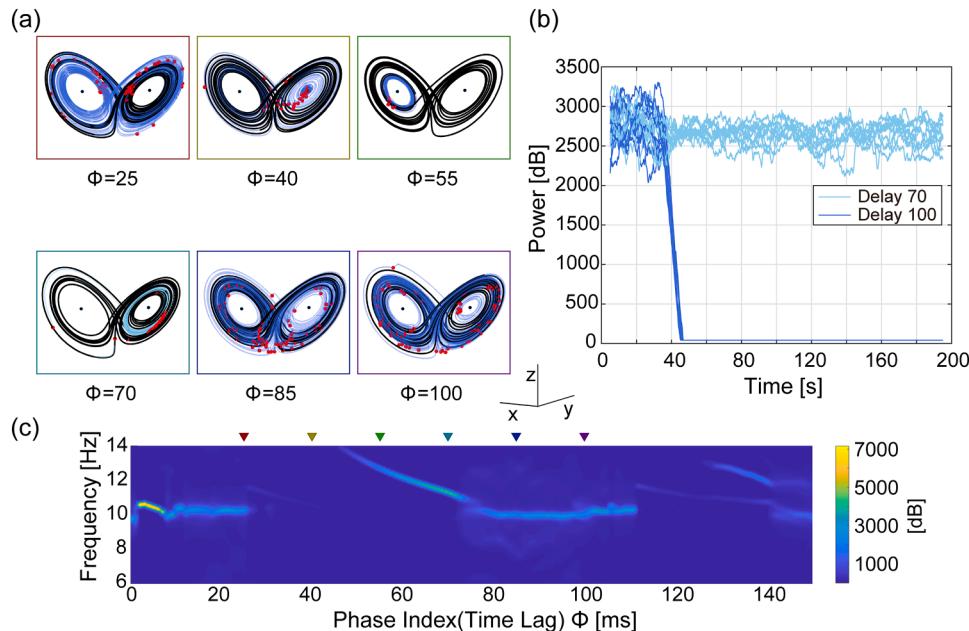


Fig. 3. Modulation results of Lorenz attractor: (a) Phase portrait of Lorenz attractor with phase index $\phi = 25, 40, 55, 70, 85, 100$ ms. Black curves in phase portraits indicate trajectories without stimulation from 38 to 40 s before the stimulus. And blue curves indicate modulated trajectories from 40 to 48 s after the stimulus. Meanwhile, the black dots indicate two equilibrium points and the red circles indicate stimulation points; (b) The power response changes over time with 10 different initial values and stimuli at the phase index $\phi = 70, 100$ ms; (c) The joint amplitude and frequency modulation effect against phase index (time lag) $\phi = 0 - 150$ ms.(For interpretation of the references to colour in this figure, the reader is referred to the web version of this article.)

complex than the fixed-point attractor. But the dynamic characteristics of each point with the same radius are still very consistent. The phase portraits of the modulation result at phase indexes $\phi = 80, 130$ are clearer, which is further investigated below in [Section 4.2](#).

For the chaotic Lorenz attractor, the simulation shows a more complex modulation result in [Fig. 3](#). It is interesting to find that it only has a

modulation effect like the online EEG modulation result under half-cycle phase index stimulation. With $\phi = 40$, the modulation effect is significant, and the oscillation behavior is greatly suppressed. With the increase of the phase index ($\phi = 55, 70$), the radius of the modulation trajectory increases as the oscillation frequency decreases. At another half-period of the phase index ($\phi = 25, 85, 100$), the modulation by PLVF

would no longer be effective.

For all four dynamic models, the difference of initial values does not influence the modulation results. To understand the modulation result of the chaotic Lorenz attractor, we analyze the dynamic characteristics near the fixed points. By setting $F_4(X) = 0$ in Eq. (8), it is found that there is only one equilibrium point with $r \leq 1$, which is at the origin $(0, 0, 0)$. With $r > 1$, two additional equilibrium points appear, which are $(\sqrt{b(r-1)}, \sqrt{b(r-1)}, r-1)$ and $(-\sqrt{b(r-1)}, -\sqrt{b(r-1)}, r-1)$. These two points are located in the center of the two wings of the butterfly trajectory. Bringing $\sigma = 80$, $\gamma = 180$, $b = 25$ into the system, we find that the eigenvalues of the local linearization matrix

$$F'_4(X) = \begin{bmatrix} -\sigma & \sigma & 0 \\ r - x_3 & -1 & -x_1 \\ x_2 & x_1 & -b \end{bmatrix} \quad (11)$$

at the two nontrivial equilibrium points are $\lambda_1 = -107.50$, $\lambda_{2,3} = 0.75 \pm 81.61i$. Hence the two nontrivial equilibrium points are unstable saddle points, but they have the same dynamic properties. Therefore, the trajectory in the system with phase index $\phi = 40, 55, 70$ for PLVF modulation may be restricted in any one wing of the butterfly. As shown in Fig. 4 with $\phi = 70$, the modulation curve converges to the left-wing, or the right-wing, which depends on the initial value. But whatever in which wing the trajectory has the same amplitude and frequency response for a given phase index.

4.2. Dynamic mechanism of PLVF modulation

To understand the joint amplitude-frequency modulation of PLVF in real EEG data, we first analyze the limit cycle model with the phase-locked stimulation at two phase indexes, $\phi = 10ms$, and $\phi = 80ms$, in Fig. 5. The simulation results of the PLVF system are shown in Fig. 6, as the phase index (time lag) ϕ increases from 0 to 200 ms.

Fig. 5a shows the dynamics of the simulated EEG rhythm modulated by an external stimulus with phase index $\phi = 10ms$. The simulated raw EEG signal is first filtered by a second-order Butterworth filter with bandwidth 8–12 Hz. Zero-crossing detection is used to recognize the phase of $3\pi/2$, and a certain time lag $\phi = 10ms$ is added to the zero-cross point to approximate the phase to be locked (upper left). The modulated EEG signal is illustrated in blue curves. Without the external stimulus, the system runs.

in a fixed orbit (black cycle at upper-right). With the external stimulus, the system reaches a stable state in less than one second, and both the power and frequency of the oscillation increase (Fig. 6b and c). As illustrated in the phase portrait (upper-right) of Fig. 5a, it is found that the external stimulus causes the radius of the trajectory to increase ($\Delta r > 0$), and the power of the oscillation correspondingly increases

($P \uparrow$). The instantaneous phase angle also increases ($\Delta\theta > 0$), as does the frequency of the oscillation ($F \uparrow$).

Fig. 5b shows the dynamics of the simulated system modulated by an external stimulus with phase index $\phi = 80ms$. The whole process of modulation is similar to that of Fig. 5a (upper-left). But the phase index ϕ increases from 10ms to 80ms. In the results, both the power and frequency of the oscillation decrease with external stimulation (Fig. 6b and c). Because the stimuli are delivered at different phases, it is found that the external stimuli decrease the trajectory radius ($\Delta r < 0$), and the power of the oscillation correspondingly decreases ($P \downarrow$). At the same time, the instantaneous phase angle decreases ($\Delta\theta < 0$), corresponding to the decreased frequency of the oscillation ($F \downarrow$).

Fig. 6 shows the joint amplitude-frequency modulation effect with phase index ϕ increasing from 0 to 200ms. The frequency decreases and shows a clear periodicity (Fig. 6b), and the power shows a sinusoidal-like shape with the increase of the phase index ϕ (Fig. 6c). The phase portrait with phase index $\phi = 10, 45, 80, 115, 150$, and 185 ms is illustrated in Fig. 6a. It is found that the increase of the trajectory radius ($\Delta r > 0$) leads to the increase of power ($P \uparrow$), and the increase of the instantaneous phase angle ($\Delta\theta > 0$) corresponds to the increase of oscillation frequency ($F \uparrow$).

4.3. Modulation result with different parameters

Considering the background noise in the EEG signal, the noise-driven fixed-point attractor model is used to explore the influence of different parameter settings on alpha modulation results. These parameters are stimulation intensity, bandwidth and center frequency in the bandpass filter, and accuracy of phase estimation.

4.3.1. Stimulation intensity

Fig. 7a shows how the modulation depth varies with stimulation intensity. The black line indicates the alpha amplitude without external stimulus, and the blue shade indicates the amplitude modulation range of the simulated alpha rhythm. The modulation depth increases with the stimulation intensity. More specifically, with the increase of stimulation intensity, the maximum amplitude of the modulated rhythm increases (upper bound of blue shadow). The minimum amplitude of the modulated rhythm (lower bound of blue shadow) first declines to reach the bottom at a stimulation intensity of 1000, and then rises with the stimulation intensity. With stimulation intensity greater than 3000, the suppression effect at certain phases in the PLVF no longer occurs. This result implies that, with the increase of stimulation intensity, the amplitude of alpha rhythm is more difficult to suppress. The frequency modulation effect shows that, with the increase of stimulation intensity, the peak frequency (bottom) is clearer and more focused, but the frequency modulation range remains unchanged (about 8–12 Hz).

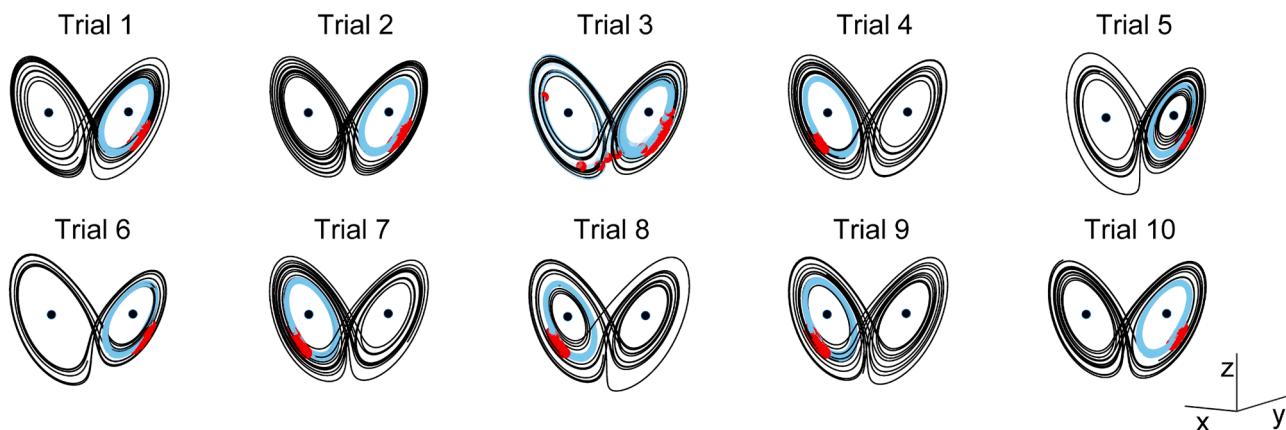


Fig. 4. The modulation results of different initial values with $\phi = 70$. The modulation curve converges to the left-wing, or the right-wing, which depends on the initial value.

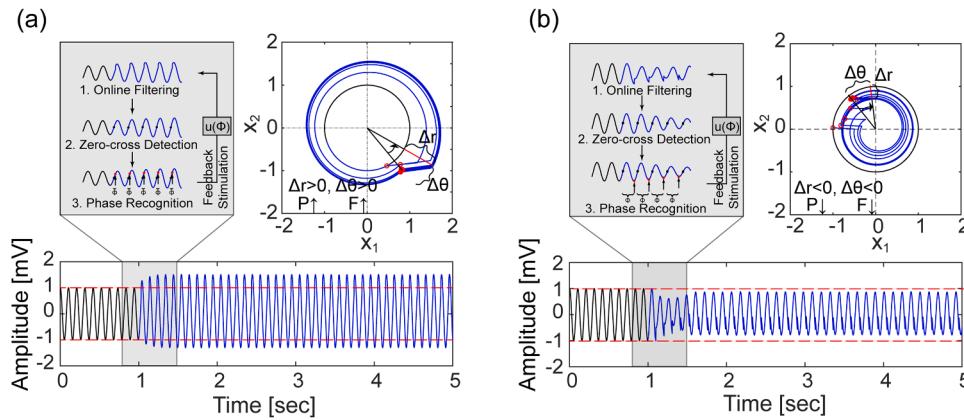


Fig. 5. Simulation results of PLVF system with external stimuli at phase index (time lag) $\phi = 10$ ms (a) and at phase index (time lag) $\phi = 80$ ms (b). Upper left in both (a) and (b): schema of output signal modulation. Black dots indicate the positive zero-crossing points. Red dots indicate the actual stimulation points. Upper right in both (a) and (b): modulation effect of the system in the phase space. Red dots indicate stimulation points. Angles of black arrows indicate angle changes of the system. Lower in both (a) and (b): output signals (simulated EEG signals). (For interpretation of the references to colour in this figure, the reader is referred to the web version of this article.)

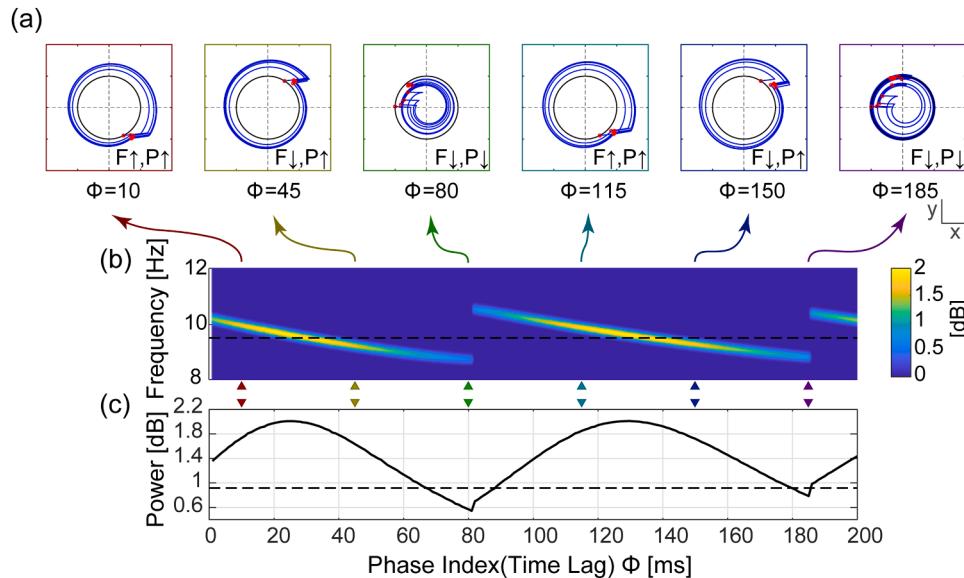


Fig. 6. Simulation results of the PLVF system with external stimuli as phase index (time lag) ϕ increases from 0 to 200 ms. (a) Modulation effect of the system in the phase space, with phase index ϕ of 10 ms, 45 ms, 80 ms, 115 ms, 150 ms, and 185 ms, respectively (left to right); (b) Power spectra of the system modulated by external stimuli delivered at different phases of ϕ ; (c) Power modulation function against phase index ϕ . Black-dashed lines indicate the frequency (b) or the power (c) without external stimuli.

4.3.2. Bandwidth of bandpass filter

Fig. 7b shows the frequency modulation range with the change of filter bandwidth. The blue shade indicates the frequency modulation range of the simulated alpha rhythm, and the blue line indicates its peak frequency. Simulation results show that the frequency modulation range narrows as the filter bandwidth decreases. The frequency modulation results (bottom) with filter bandwidths of 6–16 Hz and 10–12 Hz are similar.

4.3.3. Center frequency of bandpass filter

Fig. 7c shows the frequency modulation range with the change of center frequency of the bandpass filter. The frequency modulation range increases with the center frequency. The frequency modulation results (bottom) with filter bandwidths of 7–9 Hz and 11–13 Hz are similar.

4.3.4. Accuracy of phase estimation

Fig. 7d shows that the modulation result varies with the phase estimation error index. Added random noise is increased with the increase of the phase index to control the phase estimation error. The red line indicates the curve of exponential fitting, which is fitted by the amplitude of alpha rhythm (black line) or the attenuation coefficient (black points). The absolute value of the attenuation coefficient increases with the phase estimation error index, which means the error of modulation increases with the phase estimation error index. The peak frequency gradually becomes ambiguous with the increase of the phase estimation

error index (from 0 to 0.5 at the bottom) and phase index (0–300 ms).

5. Discussion and conclusion

In this work, we investigated the dynamic mechanism of alpha rhythm via the simulation of the proposed PLVF protocol by Huang et al. (2019). All these investigations are conducive to our deeper understanding of the dynamic structure and modulation mechanism of alpha dynamics modulation in PLVF, which can have practical significance for the modulation of alpha oscillation and further be used for potentially be used cognitive enhancement and mental diseases treatment. In the following, we discuss the three questions raised in the section of introduction and provide our answers based on the results presented in this study.

5.1. What kind of alpha dynamic model exhibits a modulation phenomenon of PLVF?

In the proposing of PLVF, Huang et al. assumed the alpha oscillation as a simple pendulum (Huang et al., 2019). However, a simple pendulum is not a stable attractor, the joint amplitude-frequency modulation result was also not as expected. In this work, we firstly investigate what kind of alpha dynamic model would exhibit a similar modulation result. This work does not aim to seek a precise model in the simulation of alpha oscillation but to qualitatively explain what dynamic structure can

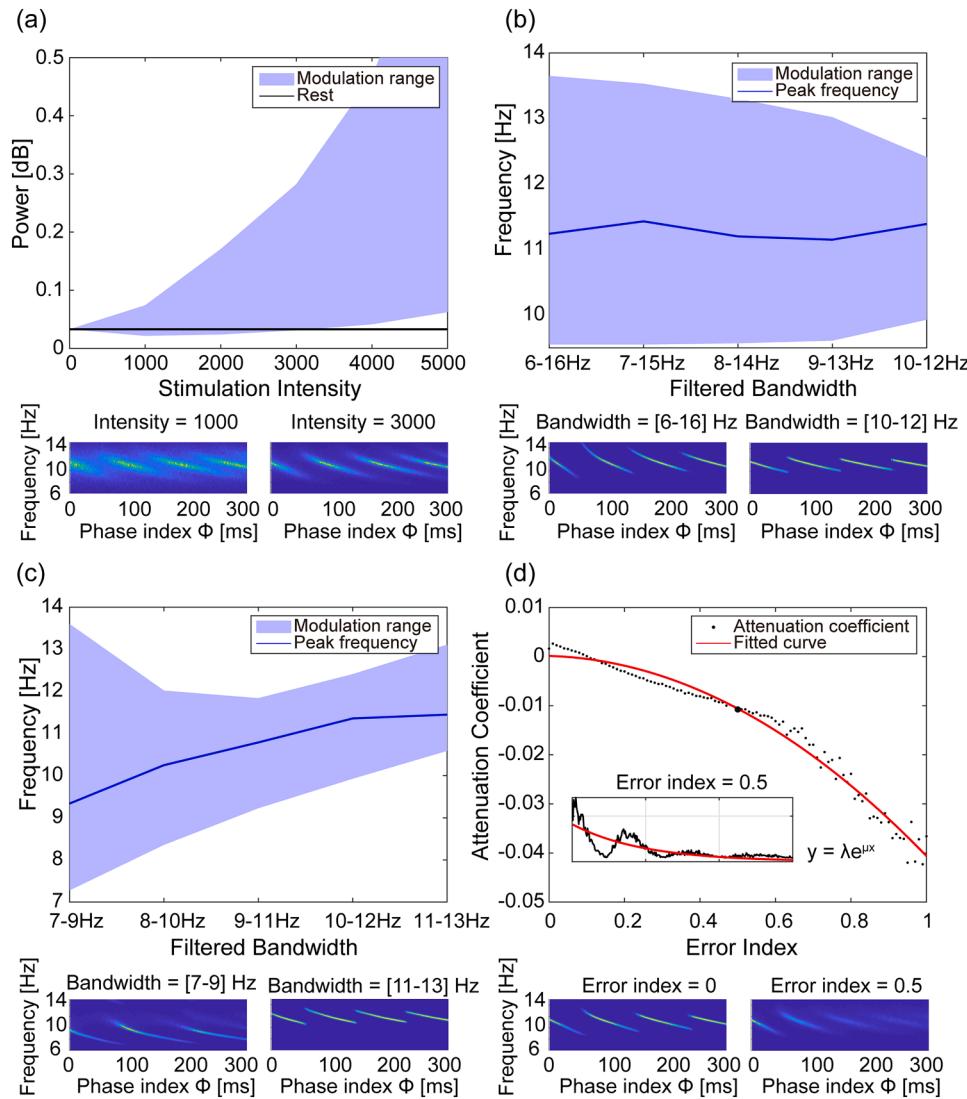


Fig. 7. Modulation results are affected by different factors: (a) stimulation intensity, (b) bandwidth of the bandpass filter, (c) center frequency of the bandpass filter, and (d) accuracy of phase estimation. (a) Upper: the result of modulation depth with the increase of stimulation intensity; Lower: frequency modulation results against phase index ϕ with a stimulation intensity of 1000 or 3000. (b) Upper: the result of frequency modulation with the decrease of bandwidth of the bandpass filter (center frequency fixed at 11 Hz); Lower: frequency modulation result against phase index ϕ with filter bandwidth of 6–16 Hz or 10–12 Hz. (c) Upper: the result of frequency modulation with the change of center frequency of the bandpass filter (bandwidth fixed at 2 Hz); Lower: frequency modulation result against phase index ϕ with filter bandwidth of 7–9 Hz or 11–13 Hz. (d) Upper: attenuation index of amplitude modulation with the increase of error of phase estimation; Lower: frequency modulation result against phase index ϕ with error index of 0 or 1.

produce such a modulation result by PLVF. Compare with the existing model, the actual alpha rhythm would be much more complex. The emphasis of mathematical modeling is not to consider all the details of the model but to abstract the core factors for the complex brain activity. Considering the large intra- and inter-subject variability, we did not make a precise estimation for the internal parameters in the model, but used the simplest models for all the three kinds of attractors to explore what kind of dynamic model exhibits a modulation phenomenon of PLVF?

Hence, the commonly used neural mass model and three well-studied attractors, including fixed-point attractor, limit-cycle attractor, and Lorenz strange attractor, were applied to simulate alpha oscillations and phase-locked feedback stimulation for alpha rhythm modulation. The simulation results show that the fixed-point attractor and the limit-cycle attractor are more consistent with the characteristics of real-world alpha dynamics. Both models have good global consistency, in that all points at different phases with the same radius have similar dynamic characteristics. They converge to a fixed point or a perfect circle by rotating at a fixed angular velocity. With the increase of the phase index ϕ , the peak frequency of alpha rhythm decreases and has clear periodicity, and the power shows a sinusoidal-like shape. On the contrary, the strong nonlinearity in the neural mass model and Lorenz attractor would destroy this global consistency, which makes the amplitude and frequency responses of the PLVF simulation different from the real EEG

result.

5.2. What is the dynamic mechanism of PLVF for alpha modulation?

Based on the simple pendulum assumption, Huang et al. (2019) proposed the closed-loop PLVF method for the modulation of alpha oscillation. Entrained by the exogenous stimulus, a joint amplitude-frequency modulation was observed from the endogenous alpha oscillations. The amplitude modulation is consistent with the expectation of the model assumption, while the peak frequency modulation does not meet the expectation. Considering the real alpha dynamic is still somewhat a black box for researchers, the experiment result on the real EEG modulation may not promote the understanding of the dynamic mechanism of PLVF-based alpha modulation. Instead, a numerical simulation based on existing experimental results can help us further understand the dynamic mechanism for alpha modulation. Based on the simulation of the limit cycle model, we find that external stimulation at a specific phase can cause the stimulated EEG signal to stabilize on a new trajectory, which leads to the modulation of amplitude. More specifically, the change of the trajectory radius Δr modulated by external stimulation causes the modulation of alpha amplitude. With a constant angular velocity in the PLVF system, the change of instantaneous phase angle $\Delta\theta$, which is stimulated by external stimulation, gives rise to the change of peak frequency. As a result, the simultaneous

modulation of trajectory radius and instantaneous phase angle leads to the joint amplitude-frequency modulation of the alpha rhythm.

5.3. Which factors affect the modulation effects in PLVF?

The influences of stimulation intensity, bandwidth, and center frequency of the bandpass filter, and accuracy of phase estimation on the modulation result are investigated. The modulation effect of the simulated alpha rhythm is affected by the stimulus intensity and the filter bandwidth. More precisely, the modulation effect on alpha amplitude is mainly affected by stimulus intensity, while the modulation effect on peak frequency is mainly influenced by filter bandwidth. It is worth noting that the suppression effect of the amplitude would disappear when the stimulus intensity is increased to a certain extent. The phenomenon can be explained by the limit-cycle attractor. When the change of trajectory radius Δr caused by external stimuli is greater than the radius of the initial trajectory, external stimuli only causes a promoting effect on the amplitude of the alpha rhythm (in Fig. 7a, the promoting effect is indicated by the blue shade above the black line, and the suppression effect is indicated by the blue shade below the black line). In addition, the accuracy of phase estimation directly affects that of the modulation result, which means the accuracy of phase estimation is fundamental to EEG modulation.

In conclusion, the results of this study provide us with greater insights into alpha rhythm modulation. By comparing four types of dynamic models in the simulation of PLVF modulation, we find that the dynamic evolution of alpha rhythm may follow a simpler dynamic structure (a fixed-point attractor or a limit-cycle attractor with globally consistent dynamic characteristics). In contrast to the highly complex functions of the brain, this result indicates that the dynamic structure of alpha rhythm may be much simpler than our expected, which even simpler than the greatly simplified lumped neural mass model. Further simulation results also explain the dynamic mechanism of PLVF for amplitude and frequency modulation of alpha rhythm and how did the parameters in the model simulation influence the modulation result. All these investigations are not only helpful for the in-depth understanding of alpha oscillation and the dynamic mechanism of PLVF modulation, but also provide theoretical guidance for the development of a precise, effective, and controllable neuromodulation technology. More specifically, the result of this study suggests that to achieve the goal of increasing or decreasing alpha rhythm, different modulation strategies should be designed. Increasing the alpha amplitude can be simply achieved by increasing the stimulation intensity in PLVF modulation, while the suppression of the alpha amplitude is more difficult, which depends on the phase of stimulation at certain intervals of the stimulation intensity.

For future work, it should be noticed that the investigation of the EEG dynamics is limited on the alpha oscillation. Currently, the modulation with PLVF methods has only been verified on the alpha band oscillation by the visual stimulus on the occipital area. Several meaningful topics need to be explored.

- Other sensory stimuli.** Based on the evidence for steady-state visual/ somatosensory/ auditory evoked potential (SSVEP/ SSSEP/ SSAEP), the visual, vibrotactile and auditory stimulus would have the largest response from their corresponding primary sensory cortex, at the frequency band around 10 Hz, 20 Hz, and 40 Hz correspondingly (Northoff et al., 2010). Whether PLVF modulation would be extended to the other band, other brain areas by other types of stimulus, and how about the neural dynamic for the modulation would be interesting to explore.

- Other stimulation techniques.** PLVF modulates the brain rhythm along the sensory pathway. Unlike the other noninvasive brain stimulation technique, like TMS, TDCS, and TACS, PLVF could not effectively work on any target brain area. But artifact-free is the most significant characteristic for PLVF modulation. Based on this, closed-

looped modulation is easy to develop by decoding the brain signal in real-time. The large artifact in TMS, TDCS, and TACS, by contrast, would make the closed-looped modulation difficult or indirectly (Noury and Siegel, 2017; Noury et al., 2016), which is considered as one of the main reasons for inter-subject variability (Frohlich and Townsend, 2021).

- Other brain rhythms.** The current work focuses on the alpha rhythm modulation with eyes open. In 2020, Philipp et al. (Philipp, 2018) reported a consistent result in the modulation of alpha rhythm with eyes closed, which demonstrates effective of the PLVF method on close eye alpha rhythm modulation. Whether the result of alpha dynamic analysis still works on other sub-bands, especially the gamma band, still needs to be verified. Gamma rhythm may have very different functions and dynamics from alpha rhythm (Herrmann and Demiralp, 2005; Le Van Quyen and Bragin, 2007). The cross-frequency coupling between gamma and theta oscillation might play a functional role in inter-cortical communication computation, and learning (Canolty and Knight, 2010; Jensen and Colgin, 2007). Recent studies show the links between gamma and cognitive functions, such as emotion (Aydin et al., 2016; Aydin, 2020), working memory (Howard et al., 2003). However, the low amplitude and high oscillation of the gamma rhythm make it difficult to be modulated by phase-locked feedback stimulus.
- Other brain functions.** The proposed PLVF could modulate the alpha oscillation on the primary sensory cortex with eyes open. How to further modulate the cognitive related brain rhythm and further affect users' cognition and behavior need to be further explored, which should be rigorously examined by well-designed experiments, large-scale validation, randomized trials, and longitudinal study, and be compared with other types of mainstream and advanced brain stimulation techniques.

Acknowledgments

This work was supported by the Science, Technology, and Innovation Commission of Shenzhen Municipality Technology Fund (No. JCYJ20190808173819182, JCYJ20170818093322718), the Shenzhen Science and Technology Program (No. JSGG20210713091811038), and the National Natural Science Foundation of China (No. 81871443). None of the authors has potential conflicts of interest to be disclosed.

Declaration of Interest Statement

None.

References

- Aydin, S., 2020. Deep learning classification of neuro-emotional phase domain complexity levels induced by affective video film clips. *IEEE J. Biomed. Health Inform.* 24, 1695–1702.
- Aydin, S., Demirtas, S., Ates, K., Tunga, M.A., 2016. Emotion recognition with eigen features of frequency band activities embedded in induced brain oscillations mediated by affective pictures. *Int. J. Neural Syst.* 26, 1650013.
- Beliaeva, V., Savvateev, I., Zerbi, V., Polania, R., 2021. Toward integrative approaches to study the causal role of neural oscillations via transcranial electrical stimulation. *Nat. Commun.* 12, 2243.
- Bollimunta, A., Mo, J., Schroeder, C.E., Ding, M., 2011. Neuronal mechanisms and attentional modulation of corticothalamic alpha oscillations. *J. Neurosci.* 31, 4935–4943.
- Brandt, M.E., 1997. Visual and auditory evoked phase resetting of the alpha EEG. *Int. J. Psychophysiol.* 26, 285–298.
- Butnik, S.M., 2005. Neurofeedback in adolescents and adults with attention deficit hyperactivity disorder. *J. Clin. Psychol.* 61, 621–625.
- Canolty, R.T., Knight, R.T., 2010. The functional role of cross-frequency coupling. *Trends Cogn. Sci.* 14, 506–515.
- Choi, S.W., Chi, S.E., Chung, S.Y., Kim, J.W., Ahn, C.Y., Kim, H.T., 2011. Is alpha wave neurofeedback effective with randomized clinical trials in depression? a pilot study. *Neuropsychobiology* 63, 43–51.
- Cohen, M.X., 2017. Where does EEG come from and what does it mean? *Trends Neurosci.* 40, 208–218.
- Debener, S., Ullsperger, M., Siegel, M., Engel, A.K., 2006. Single-trial EEG-fMRI reveals the dynamics of cognitive function. *Trends Cogn. Sci.* 10, 558–563.

- Fox, D.J., Tharp, D.F., Fox, L.C., 2005. Neurofeedback: an alternative and efficacious treatment for attention deficit hyperactivity disorder. *Appl. Psychophysiol. Biofeedback* 30, 365–373.
- Frohlich, F., Townsend, L., 2021. Closed-loop transcranial alternating current stimulation: towards personalized non-invasive brain stimulation for the treatment of psychiatric illnesses. *Curr. Behav. Neurosci. Rep.* 8, 51–57.
- Glass, L., Kaplan, D.T., Lewis, J.E., 1993. Tests for deterministic dynamics in real and model neural networks. In: Proceedings of the 2nd Annual Conference on Nonlinear Dynamics Analysis of the EEG, World Scientific, Singapore. pp. 223–249.
- Grandy, T.H., Werkle-Bergner, M., Chicherio, C., Schmiedek, F., Loevden, M., Lindenberger, U., 2013. Peak individual alpha frequency qualifies as a stable neurophysiological trait marker in healthy younger and older adults. *Psychophysiology* 50, 570–582.
- Grimbert, F., Faugeras, O., 2006. Bifurcation analysis of Jansen's neural mass model. *Neural Comput.* 18, 3052–3068.
- Halgren, M., Ulbert, I., Bastuji, H., Fabó, D., Eröss, L., Rey, M., Devinsky, O., Doyle, W.K., Mak-McCully, R., Halgren, E., Wittner, L., Chauvel, P., Heit, G., Eskandar, E., Mandell, A., Cash, S.S., 2019. The generation and propagation of the human alpha rhythm. *Proc. Natl. Acad. Sci.* 116, 23772–23782.
- Hammond, D.C., 2005. Neurofeedback treatment of depression and anxiety. *J. Adult Dev.* 12, 131–137.
- Hanslmayr, S., Sauseng, P., Doppelmayr, M., Schabus, M., Klimesch, W., 2005. Increasing individual upper alpha power by neurofeedback improves cognitive performance in human subjects. *Appl. Psychophysiol. Biofeedback* 30, 1–10.
- Hanslmayr, S., Gross, J., Klimesch, W., Shapiro, K.L., 2011. The role of alpha oscillations in temporal attention. *Brain Res. Rev.* 67, 331–343.
- Herrmann, C.S., Demiralp, T., 2005. Human EEG gamma oscillations in neuropsychiatric disorders. *Clin. Neurophysiol.* 116, 2719–2733.
- Howard, M.W., Rizzuto, D.S., Caplan, J.B., Madsen, J.R., Lisman, J., Aschenbrenner-Scheibe, R., Schulze-Bonhage, A., Kahana, M.J., 2003. Gamma oscillations correlate with working memory load in humans. *Cereb. Cortex* 13, 1369–1374.
- Hsueh, J.J., 2017. Neurofeedback training of EEG alpha rhythm enhances episodic and working memory (vol 37, pg 2662, 2016). *Hum. Brain Mapp.* 38, 3315.
- Huang, G., Zhang, D., Meng, J., Zhu, X., 2011. Interactions between two neural populations: a mechanism of chaos and oscillation in neural mass model. *Neurocomputing* 74, 1026–1034.
- Huang, G., Liu, J., Li, L., Zhang, L., Zeng, Y., Ren, L., Ye, S., Zhang, Z., 2019. A novel training-free externally-regulated neurofeedback (ER-NF) system using phase-guided visual stimulation for alpha modulation. *Neuroimage* 189, 688–699.
- Ince, R., Adanir, S.S., Sevmez, F., 2020. The inventor of electroencephalography (EEG). *Childs Nervous System*. Hans Berger, pp. 1873–1941.
- Jansen, B.H., Rit, V.G., 1995. Electroencephalogram and visual evoked potential generation in a mathematical model of coupled cortical columns. *Biol. Cybern.* 73, 357–366.
- Jansen, B.H., Zouridakis, G., Brandt, M.E., 1993. A neurophysiologically-based mathematical model of flash visual evoked potentials. *Biol. Cybern.* 68, 275–283.
- Jensen, O., Colgin, L.L., 2007. Cross-frequency coupling between neuronal oscillations. *Trends Cogn. Sci.* 11, 267–269.
- Kasten, F.H., Herrmann, C.S., 2019. Recovering brain dynamics during concurrent tACS-M/EEG: an overview of analysis approaches and their methodological and interpretational pitfalls. *Brain Topogr.* 32, 1013–1019.
- Kirschfeld, K., 2005. The physical basis of alpha waves in the electroencephalogram and the origin of the “Berger effect”. *Biol. Cybern.* 92, 177–185.
- Le Van Quyen, M., Bragin, A., 2007. Analysis of dynamic brain oscillations: methodological advances. *Trends Neurosci.* 30, 365–373.
- Lopes da Silva, F.H., Vos, J.E., Mooibroek, J., Vanrotterdam, A., 1980. Relative contributions of intracortical and thalamo-cortical processes in the generation of alpha rhythms, revealed by partial coherence analysis. *Electroencephalogr. Clin. Neurophysiol.* 50, 449–456.
- Lopes da Silva, F.H., Pijn, J.P., Velis, D., Nijssen, P.C.G., 1997. Alpha rhythms: noise, dynamics and models. *Int. J. Psychophysiol.* 26, 237–249.
- Maltseva, I.V., Masllobœuf, Y.P., 1997. Alpha rhythm parameters and short-term memory span. *Int. J. Psychophysiol.* 27, 169–180.
- Mansouri, F., Fettes, P., Schulze, L., Giacobbe, P., Zariffa, J., Downar, J., 2018. A real-time phase-locking system for non-invasive brain stimulation. *Front. Neurosci.* 12, 877.
- Northoff, G., Qin, P.M., Nakao, T., 2010. Rest-stimulus interaction in the brain: a review. *Trends Neurosci.* 33, 277–284.
- Noury, N., Siegel, M., 2017. Phase properties of transcranial electrical stimulation artifacts in electrophysiological recordings. *Neuroimage* 158, 406–416.
- Noury, N., Hipp, J.F., Siegel, M., 2016. Physiological processes non-linearly affect electrophysiological recordings during transcranial electric stimulation. *Neuroimage* 140, 99–109.
- Palus, M., 1993. Testing for nonlinearity in the EEG. In: Proceedings of the 2nd Annual Conference on Nonlinear Dynamical Analysis of the EEG, World Scientific, Singapore. pp. 100–14.
- Pereda, E., Quiroga, R.Q., Bhattacharya, J., 2005. Nonlinear multivariate analysis of neurophysiological signals. *Prog. Neurobiol.* 77, 1–37.
- Philipp, S., 2018. Closed-loop photic stimulation using EEG. *Neuropsychobiology* 77, 155.
- Schreckenberger, M., Lange-Asschenfeldt, C., Lochmann, M., Mann, K., Siessmeier, T., Buchholz, H.G., Bartenstein, P., Gründer, G., 2006. The thalamus as the generator and modulator of EEG alpha rhythm: a combined PET/EEG study with lorazepam challenge in humans. *Neuroimage* 32, 485.
- Schwab, K., Ligges, C., Jungmann, T., Hilgenfeld, B., Haueisen, J., Witte, H., 2006. Alpha entrainment in human electroencephalogram and magnetoencephalogram recordings. *Neuroreport* 17, 1829–1833.
- Stam, C.J., Pijn, J.P.M., Suffczynski, P., Lopes da Silva, F.H., 1999. Dynamics of the human alpha rhythm: evidence for non-linearity? *Clin. Neurophysiol.* 110, 1801–1813.
- Vijayan, S., Kopell, N.J., 2012. Thalamic model of awake alpha oscillations and implications for stimulus processing. *Proc. Natl. Acad. Sci.* 109, 18553–18558.
- Zarubin, G., Gundlach, C., Nikulin, V., Villringer, A., Bogdan, M., 2020. Transient amplitude modulation of alpha-band oscillations by short-time intermittent closed-loop tACS. *Front. Hum. Neurosci.* 14, 366.
- Zhang, L., 2017. Artificial neural networks model design of Lorenz chaotic system for EEG pattern recognition and prediction. In: Proceeding of IEEE Life Science Conference. pp. 39–42.