

RESEARCH ARTICLE | JUNE 12 2024

Decoding electroencephalographic responses to visual stimuli compatible with electrical stimulation

Simone Romeni   ; Laura Toni  ; Fiorenzo Artoni  ; Silvestro Micera 



APL Bioeng. 8, 026123 (2024)
<https://doi.org/10.1063/5.0195680>



APL Energy
Latest Articles Online!
Read Now



Decoding electroencephalographic responses to visual stimuli compatible with electrical stimulation



Cite as: APL Bioeng. 8, 026123 (2024); doi: 10.1063/5.0195680

Submitted: 3 January 2024 · Accepted: 20 May 2024 ·

Published Online: 12 June 2024



View Online



Export Citation



CrossMark

Simone Romeni,^{1,2,a)}  Laura Toni,^{1,3}  Fiorenzo Artoni,⁴  and Silvestro Micera^{1,2,3} 

AFFILIATIONS

¹Modular Implantable Neurotechnologies (MINE) Laboratory, Università Vita-Salute San Raffaele & Scuola Superiore Sant'Anna, Milan, Italy

²Bertarelli Foundation Chair in Translational Neural Engineering, Center for Neuroprosthetics and Institute of Bioengineering, Ecole Polytechnique Federale de Lausanne, Lausanne, Switzerland

³The Biorobotics Institute and Department of Excellence in Robotics and AI, Scuola Superiore Sant'Anna, Pontedera, Italy

⁴Department of Clinical Neurosciences, Faculty of Medicine, University of Geneva, Geneva, Switzerland

^{a)} Author to whom correspondence should be addressed: simone.romeni@epfl.ch

ABSTRACT

Electrical stimulation of the visual nervous system could improve the quality of life of patients affected by acquired blindness by restoring some visual sensations, but requires careful optimization of stimulation parameters to produce useful perceptions. Neural correlates of elicited perceptions could be used for fast automatic optimization, with electroencephalography as a natural choice as it can be acquired non-invasively. Nonetheless, its low signal-to-noise ratio may hinder discrimination of similar visual patterns, preventing its use in the optimization of electrical stimulation. Our work investigates for the first time the discriminability of the electroencephalographic responses to visual stimuli compatible with electrical stimulation, employing a newly acquired dataset whose stimuli encompass the concurrent variation of several features, while neuroscience research tends to study the neural correlates of single visual features. We then performed above-chance single-trial decoding of multiple features of our newly crafted visual stimuli using relatively simple machine learning algorithms. A decoding scheme employing the information from multiple stimulus presentations was implemented, substantially improving our decoding performance, suggesting that such methods should be used systematically in future applications. The significance of the present work relies in the determination of which visual features can be decoded from electroencephalographic responses to electrical stimulation-compatible stimuli and at which granularity they can be discriminated. Our methods pave the way to using electroencephalographic correlates to optimize electrical stimulation parameters, thus increasing the effectiveness of current visual neuroprostheses.

© 2024 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>). <https://doi.org/10.1063/5.0195680>

INTRODUCTION

Electrical stimulation (ES) can elicit visual sensations in a large number of patients affected by acquired blindness.^{1,2} Stimulation can be delivered via retinal^{3,4} or optic nerve^{5,6} prostheses for retinal degeneration or cortical prostheses⁷ for optic neuropathies or stroke-induced blindness. In most visual prostheses, a camera captures a stream of visual scenes, which are first processed through computer vision algorithms^{8–12} and then converted to stimulation patterns through an *a priori* model of the visual sensations (phosphenes) produced by stimulation. In retinal stimulation, it is customary to model

phosphenes as round and to build stimulation protocols from “pixelizations” of the target visual scenes,^{13,14} resulting in suboptimal stimulation protocols as this assumption is only approximately true.¹⁵ The “pixelization” approach is even more suboptimal in the case of optic nerve stimulation because of the very elongated phosphenes generated in this case.¹⁶

In Ref. 17, we employed neural network models inspired DiCarlo and colleagues' work^{18,19} to show *in silico* that the ES of the optic nerve may be optimized by the cortical activation elicited by stimulation as a feedback signal. A similar approach has been proposed *in vivo* in

mice,²⁰ with both stimulation and recordings at the same cortical area. Such “end-to-end” optimization strategies evaluate the performance of different candidate stimulation protocols implicitly taking into account the limits of current interfaces, while in other encoding techniques the construction of stimulation targets does not take them into account. Moreover, such optimization process does not require knowledge of all the anatomical and functional details necessary to build a detailed biophysical model like the ones available for animal models, which still require very invasive or destructive imaging techniques.²¹

Here, we investigate whether electroencephalography (EEG) can be used as a feedback cortical signal to optimize ES protocols, allowing stimulation protocol optimization without resorting to invasive procedures except for the neuroprosthesis implantation. EEG has been used in the past to characterize the cortical response to electrical stimulation,^{22–25} but no framework has been formulated for automatic optimization based on EEG to the best of our knowledge. Even though EEG is an ideal candidate in terms of non-invasiveness, it is characterized by poor signal-to-noise ratio (SNR) and spatial resolution,²⁶ which could prevent the discrimination of structural features of visual stimuli. In addition, in our case, we need to consider stimuli compatible with the perceptions elicited by ES (phosphenes), consisting of solid bright blobs on a uniform background. Increasing the stimulation amplitude of a given electrode contact increases the size of the solid blob and thus the average luminance of the stimulus (as it recruits more cells without substantially altering their firing rates^{16,27}). This makes our ideal visual stimuli quite different from those employed in neuroscientific settings, where investigations have focused either on the semantic valence of “complex” images, representing, e.g., human faces, animals, and everyday-life objects,^{28–31} or on the response to structural variations and in different subject cohorts of “simple” stimuli like distributed gratings^{32–36} or checkerboards.^{37–39} In general, the different visual stimuli in such datasets vary the value of a single feature at a time, e.g., the size of isoluminant stimuli. This implies that the results obtained by existing studies may not be immediately translatable to our scenario and motivated us to acquire a new dataset of EEG correlates, characterizing the response to such phosphene-like visual stimuli. We then proceeded to investigate whether the distributions of the EEG correlates with several stimulus features of our visual stimuli were statistically distinguishable and to evaluate whether a classification algorithm could reliably assign an EEG response to a given feature class.

Such EEG response classifiers could be used to “evolve” ES protocols until the elicited response is assigned to a given target class as illustrated in Fig. 1(a). The use of an EEG decoder allows optimization of stimulation parameters relying on a quantitative marker and may substantially speed up the optimization process, as also discussed in Ref. 17. While such evolution should ideally happen on the basis of the decoding of single EEG trials, it is possible that the information from multiple trials may be required because of the very low SNR of EEG signals.²⁶ Multi-trial decoding can be performed by simple concatenation of several trials along the time- or the channel-dimension or by training and testing the decoder on event related potentials (ERPs).⁴⁰ The former approach leads to results dependent upon trial ordering unless complex training schemes are employed, while the latter approach requires a very large dataset, so that an acceptable number of

samples remains after trial averaging. Here, instead, we aggregated the class probabilities for the single trials.^{40,41} We then characterized the improvement in decoding accuracy produced by using several trials at a time with respect to a single one.

RESULTS

In the present work, we analyzed data from ten healthy subjects positioned in front of a screen covering most of their visual field [Fig. 1(b)]. We flashed different rectangular bright stimuli on a black background, for a total of 50 repetitions of each one of 60 different stimuli (see [supplementary material](#) Fig. 1). Each visual stimulus was flashed in and kept for 750 ms, then flashed off. The next stimulus was flashed in after an inter-stimulus time randomly chosen in the range of (1000 and 1250) ms [Fig. 1(c)]. In order to assess the possibility to discriminate several features of a stimulus, we defined a set of decoding tasks, respectively, to decode information about the location of the presented stimulus, its size, or both. To do this, we formed classes from different sets of the 60 original stimuli, and we determined ten decoding tasks where it was necessary to discriminate whether a stimulus belonged to one of several mutually exclusive classes. The set of stimulus classes and decoding tasks is detailed in [supplementary material](#) Fig. 2.

Figure 2 shows the ERPs in the occipital channels for the bars and blocks tuning sets, to give an example of modulations related to the size of the stimuli and thus to their luminosity. In general, larger visual stimuli produce a larger positive and a smaller negative deviation in the ERP with respect to baseline. This means that scaling the size of a given visual stimulus does not result in a scaled copy of the EEG response as it could be expected. The dependency of the ERP on the stimulus size is similar across different subjects, even considering inter-subject variations in the ratios between the peak amplitudes for different stimulus sizes. Instead, Fig. 3 displays the ERPs in the occipital channels for the coarse left vs right and superior vs inferior tuning sets, to give an example of modulations related to the location of the stimuli, irrespective of their size. The inter-subject differences in the ERPs corresponding to stimulus in different locations of the visual field are more marked. For example, stimuli in the left part of the visual field produced a larger negative deflection in the right part of the scalp, and stimuli in the lower part of the visual field produced larger negative deflections in the ERPs. In both cases, the time intervals where the responses for the different classes were significantly different are shown. For most tasks and subjects, several regions reached statistical discriminability ($p < 0.05$) among the different classes.

In Fig. 4(a), subject-wise accuracies for the tenfold cross validation folds or Monte Carlo resamples are shown with the corresponding chance level. As can be seen from Fig. 4(b), the decoding accuracy is significantly above chance both in the location decoding tasks and in the blocks decoding tasks. When accuracy is not significantly higher than chance for a task and subject, it tends to be so for more than one subject, suggesting an intrinsic difficulty of the decoding task, rather than a low signal quality from the specific subject. The decoding performance is variable across subjects, confirming an important inter-subject variability. As expected, we noticed that our decoding algorithm had worst performance for tasks that included many classes whose stimuli differ only slightly (e.g., blocks and bars, where the difference between bar3, block4, and bar5 is only one block), or when concurrent increases in

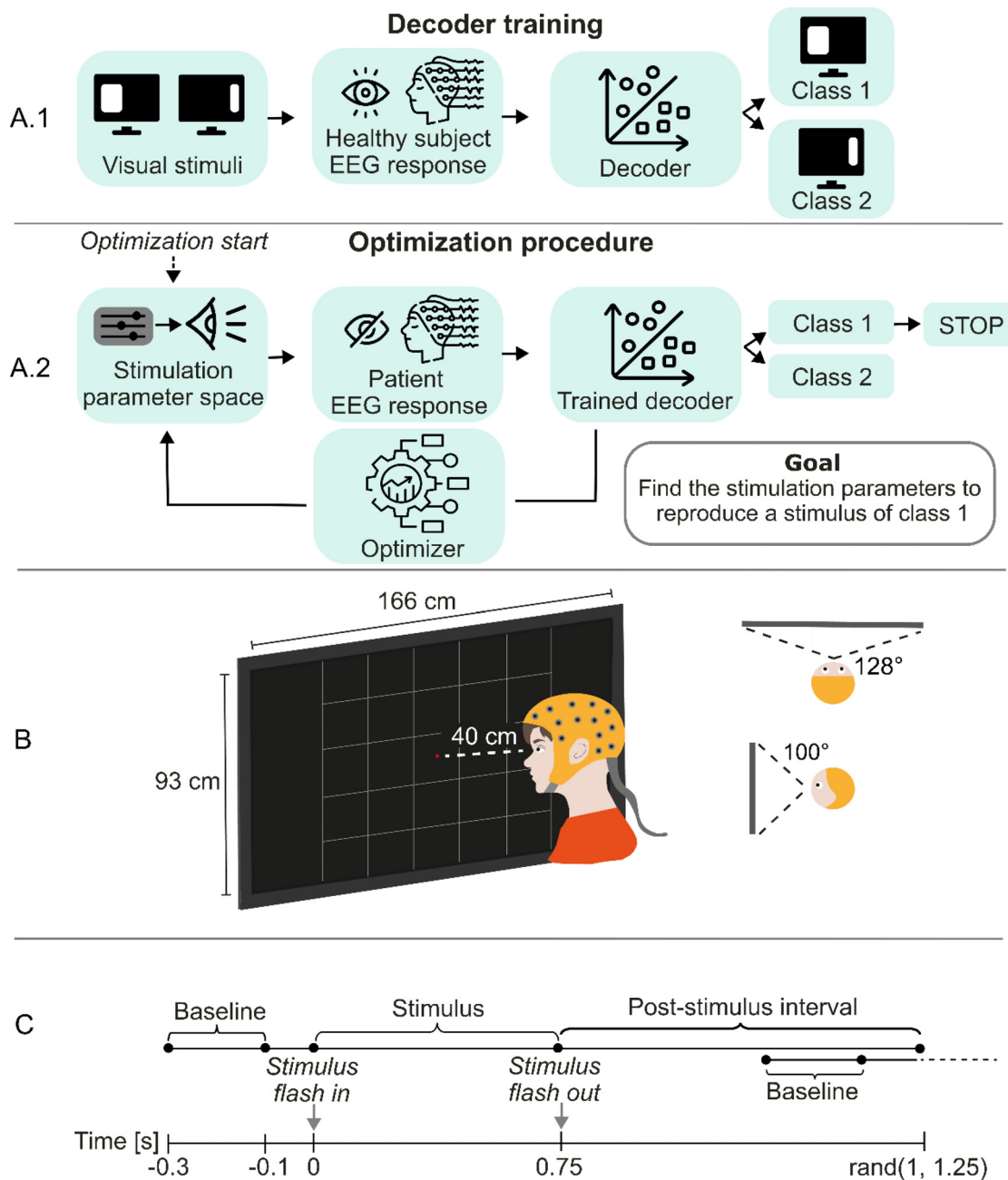


FIG. 1. Experimental setup. (a) Workflow for the use of an EEG decoder of visual features in the optimization of electrical stimulation for vision restoration. (b) Positioning of the subjects with respect to the screen, screen dimensions, and covered portion of the visual field. Stimuli were flashed on a region of the screen, divided into 25 equally sized squares. Participants were asked to fixate the center of the screen where a red dot was presented. (c) Representation of the visual stimulation sequence, consisting of baseline, stimulus, and random post-stimulus intervals.

luminosity and eccentricity of the stimulus were present. In the latter case, higher luminosity would tend to produce high ERP responses, while higher eccentricity would lead to decreased ERP amplitudes, possibly leading to a compensation of the two effects (out vs center decoding task).

Decoding scores for three subjects with and without ICA removal of ocular activity showed no substantial differences (supplementary material Fig. 3). In supplementary material Fig. 4, we show features' importance for the RBF SVM decoding for left vs right and superior vs inferior portions in subject 1. The complete

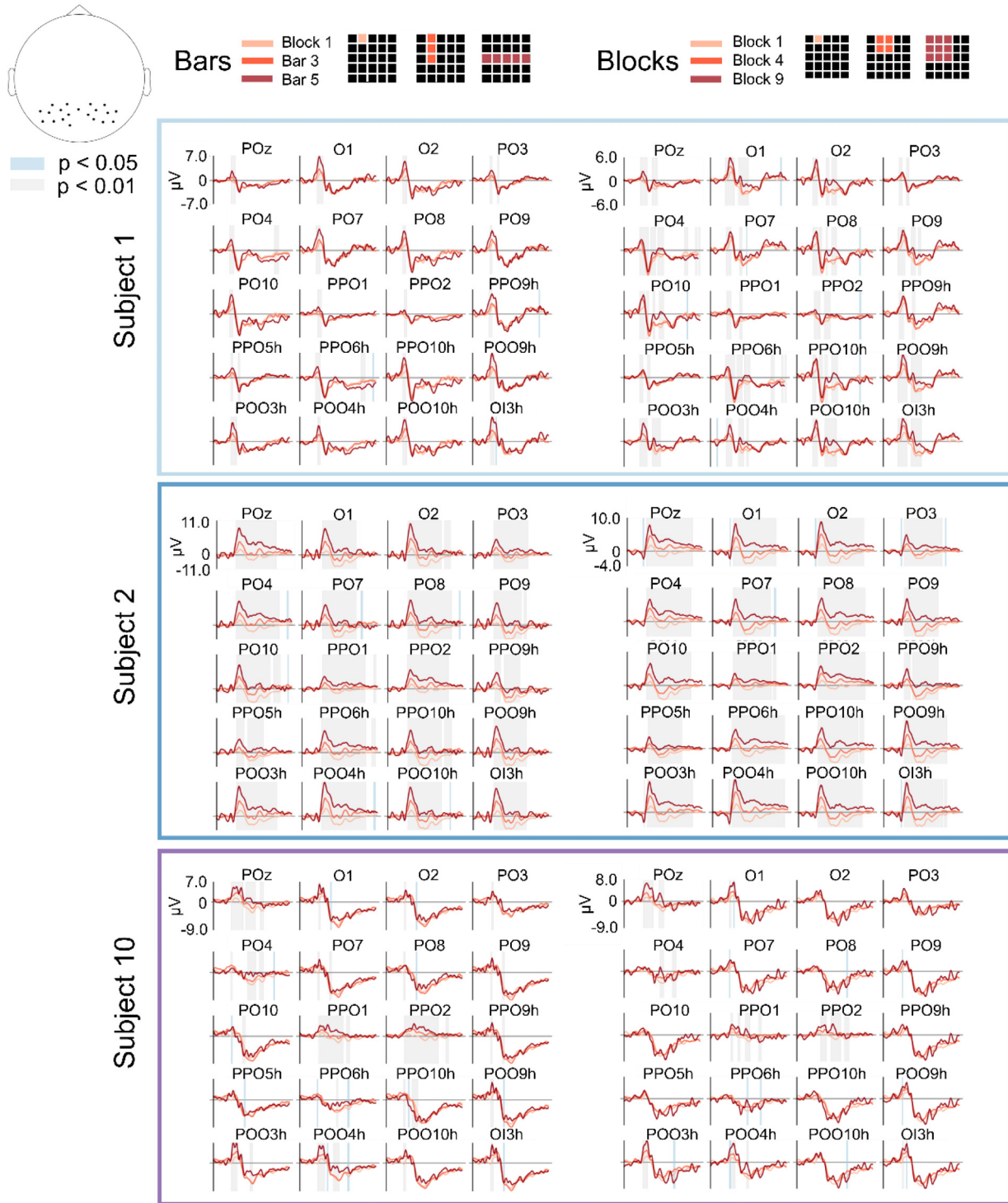


FIG. 2. ERPs for luminosity-related classes. ERPs (0–0.75 s) for luminosity-related classes (1-block vs 3-bars vs 5-bars, and 1-block vs 4-blocks vs 9-blocks) for three subjects in the occipital and parietal channels. Chosen subjects (subjects 1, 2, and 10) are those that had best, worst, and average classification scores. Different colors correspond to different stimuli, while colored bands correspond to time intervals where the single-trial responses for the different classes are significantly different, with the band color representing the level of significance. The location of the shown channels is displayed in the topographic map in the top left.

03 July 2024 09:58:52

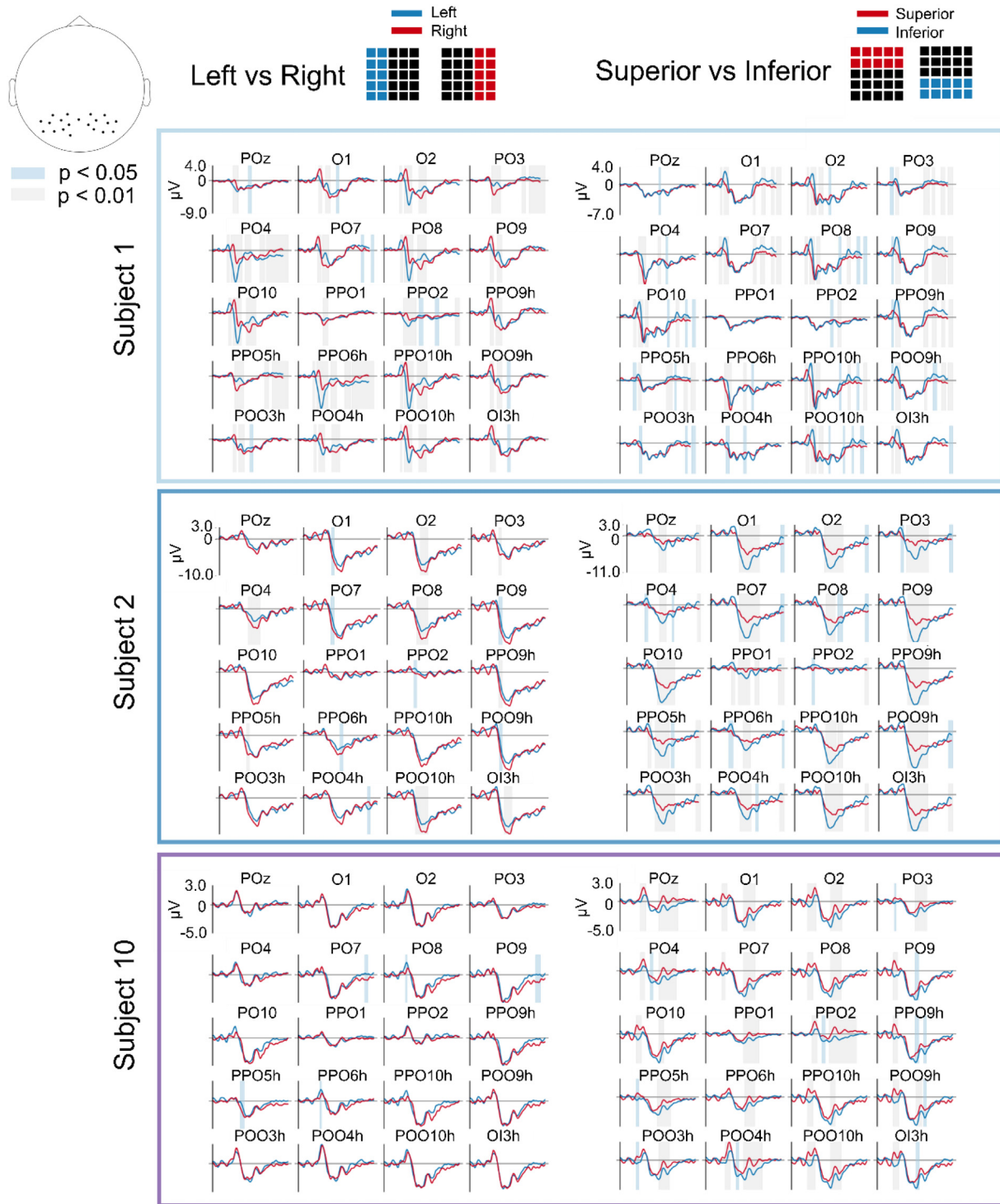


FIG. 3. ERPs potentials for location-related classes. ERPs (0–0.75 s) for location-related classes (left vs right, and superior vs inferior) for three subjects in the occipital and parietal channels. Chosen subjects (subjects 1, 2, and 10) are those that had best, worst, and average classification scores. Different colors correspond to different stimuli, while colored bands correspond to time intervals where the single-trial responses for the different classes are significantly different, with the band color representing the level of significance (see legend). The location of the shown channels is displayed in the topographic map at the top left corner.

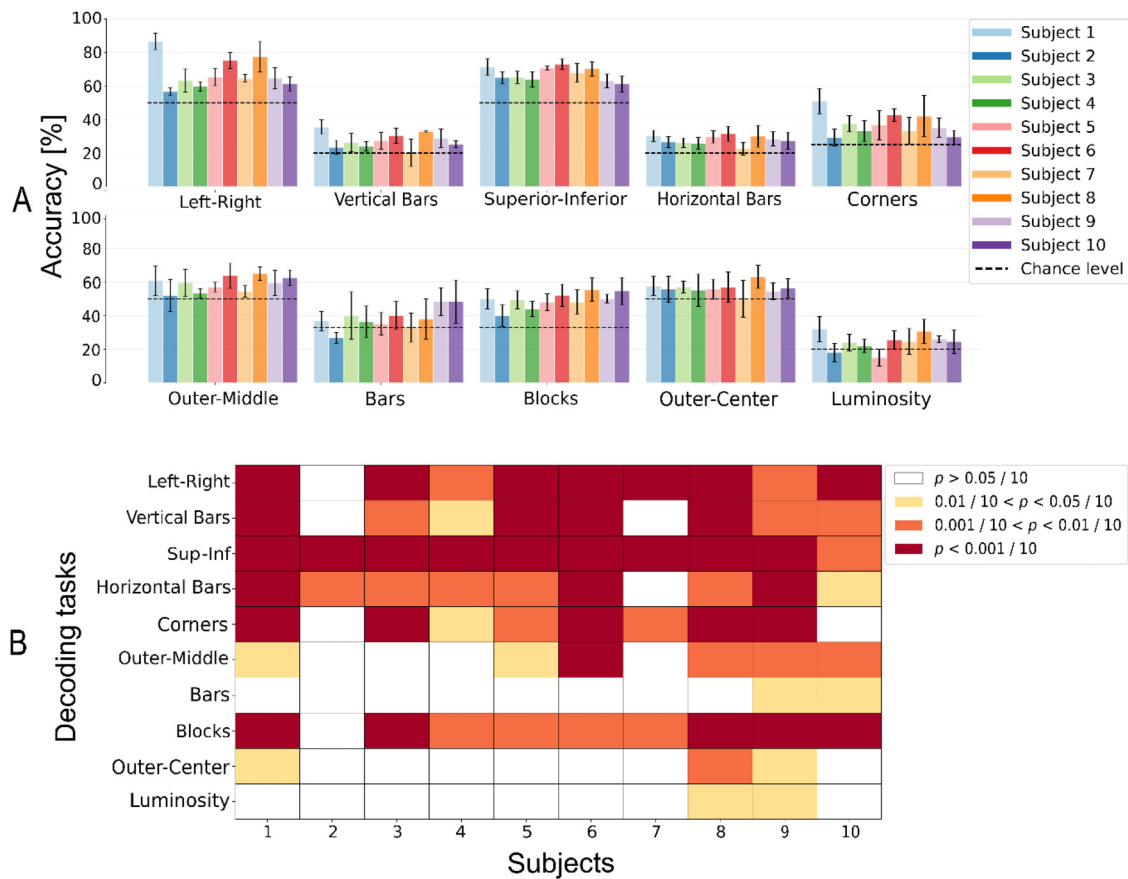


FIG. 4. Single-trial decoding. (a) Classification accuracy for all subjects and decoding tasks, evaluated as the mean of a tenfold cross-validation or ten resamplings. Error bars represent the standard deviation over the ten folds and resamplings. (b) P-values obtained from the Wilcoxon test for all macro-classes and all subjects. The analysis tests if the median accuracy obtained with single-trial decoding is greater than that obtained with a random decoder. Cells have been highlighted in red if $p < 0.0001$, orange if $p < 0.001$, yellow if $p < 0.005$, and white if $p > 0.005$.

map of features' importance and averaging across time to get a spatial map and across channels to get a temporal map of importance are displayed.

In Fig. 5(a), it is possible to observe how the performance of multi-trial decoding varies when an increasing number of trials is employed in the decoding. Multi-trial decoding can lead to important increases in accuracy, even if different subjects lead to different performance increases. Interestingly, in some cases multi-trial decoding led to a decrease in accuracy. To investigate such phenomenon, the performance for single- and multi-trials for each cross-validation fold or Monte Carlo resampling is shown in [supplementary material](#) Fig. 5, together with the confidence interval of the random decoder for the given task and subject. We then investigated whether there is a dependency of the accuracy variation led by multi-trial decoding upon the initial accuracy and show the results of this analysis in Fig. 5(b). When such dependency was statistically significant, we found a positive correlation between the two variables, meaning that multi-trial decoding produced higher performance increases with respect to single-trial decoding when the initial performance of the single-trial was higher.

DISCUSSION

In this work, we defined a novel set of visual stimuli compatible with ES of the visual system and showed that the EEG ERPs corresponding to different stimulus classes are statistically distinguishable. We then performed single-trial decoding using traditional ML, obtaining above chance-level accuracies, and improved them with a multi-trial decoding strategy.

The ability to decode visual stimuli from EEG activity is a premise for the optimization of the ES using such technique as feedback. In fact, it can allow determination of whether the applied electrical stimulation produces a sensation with specific target features without the need to rely on subjective feedback collected via questionnaires. The optimization process exploits a set of decoders trained on healthy subjects, where the association between visual perception and EEG correlate is known, as it corresponds to the visual stimulus generating the EEG response. The optimization loop would thus consist in choosing a target perception, applying a stimulation protocol on the patient, characterizing the features of the elicited perception passing the EEG evoked activity through the above-mentioned decoders (establishing

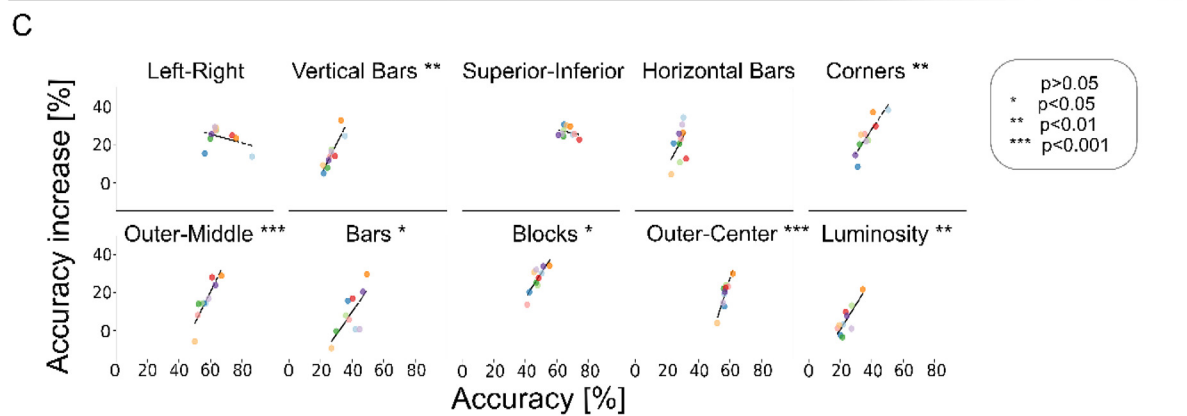
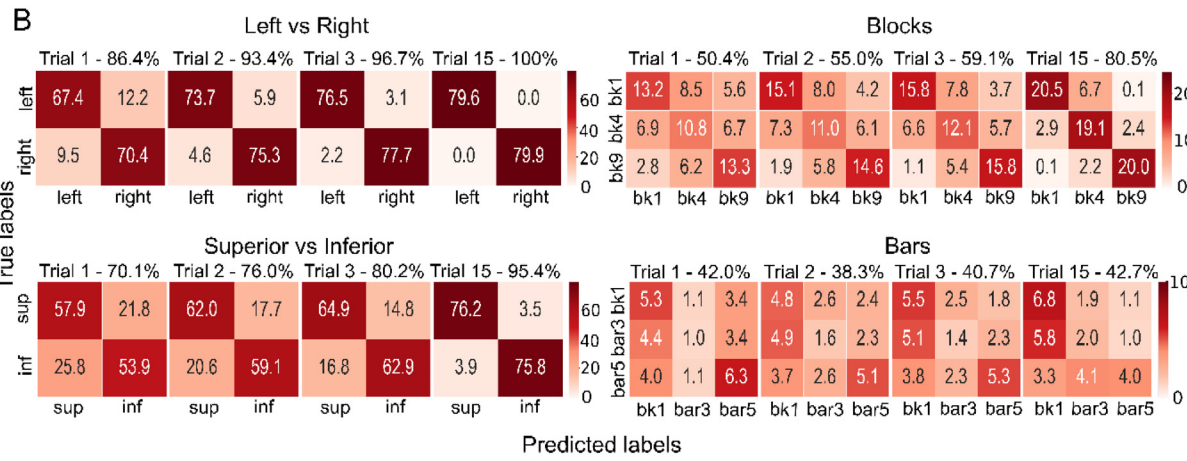
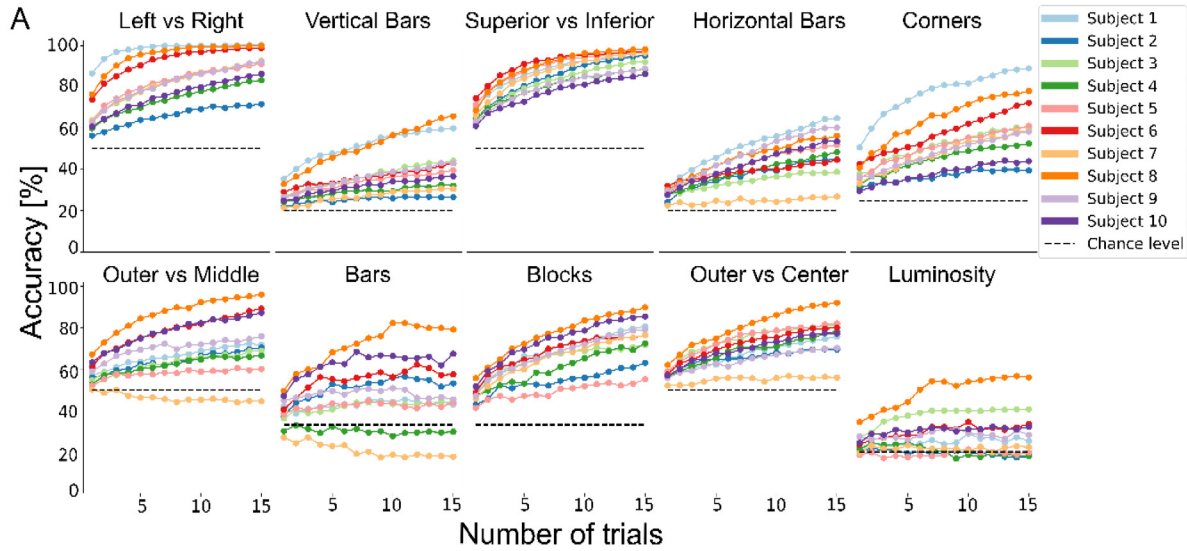


FIG. 5. Multi-trial decoding results. (a) Subject-wise accuracy curves obtained by varying the number of trials used in the multi-trial inference. (b) Dependency of the variation of accuracy from single-trial to 15-trial (multi-trial) inference upon the single-trial variation. Positive values account for an increase in accuracy between single- and multi-trial inference. The black line represents the least squares fit of the points in each subplot, and the levels of significance associated with the line slopes are reported in the subplot titles. (c) Average confusion matrices for subject 1, single-, 2-, 3-, and 15-trial inference. Averages are computed across the tenfold cross validation/10 Monte Carlo resamplings. In the subplot titles, we report the average accuracy across folds.

whether the evoked perception occupies the left or right, superior or inferior visual field), and proposing the next candidate stimulation to minimize the discrepancy between the features of the evoked and target perceptions.

In our analyses, preprocessing was kept to a minimum, i.e., to bad channel rejection and frequency filtering. This is coherent with our final goal of using such analyses to optimize visual stimulation in real-time. Indeed, bad channel rejection can be performed at the beginning of acquisition. During the optimization experiments, after each candidate stimulus is administered, the corresponding EEG epoch is filtered (which can be done in real-time) and passed through a set of decoders establishing its similarity to a given target natural stimulus so that the next candidate stimulus is determined. Instead, the application of other state-of-the-art methods to improve signal quality like, e.g., ICA can be time intensive and normally requires expert input (e.g., manually selecting artefactual independent components), which would add a substantial amount of time between the presentation of different candidate stimuli, reducing the amount of explorable candidates during an experimental session. Additionally, recent works report that the use of such techniques often leads to marginal or no improvement in the target application,⁴² thus not justifying a drastic reduction in the throughput of our optimization routine.

The presented ERPs do not show the typical peaks studied in the literature related to visual responses like N70 and P100. It is interesting to remark that even though not visible in the ERPs, the information present at such latencies is exploited by the employed RBF SVMs, as we observed peaks of feature importance in those locations, particularly evident in occipital and parietal areas, corresponding to visual sensory areas. Thus, the peaks could be covered by the surrounding neural activity and not be visible through averaging but they may be highlighted by the nonlinear processing performed by our kernel machine. The very small amplitude of the components linked to lower sensory cortical areas may be linked to the fact that they are generally observed in correspondence to variations in stimulus spatial frequency and orientation.⁴³ Additionally, it is known that stimuli of large size trigger surround suppression in lower visual areas because of the tendency of neurons to suppress the activation of other neurons with neighboring receptive fields.^{44–46} Responses with latencies similar to ours have been found in healthy controls under visual stimulation in Ref. 47.

Typically, EEG decoding is performed either by extracting features defined in the literature and passing them to traditional machine learning or by passing the whole spatiotemporal signal to deep learning (DL) architectures. In our case, neither strategy was feasible. The lack of knowledge about the present class of stimuli prevented us from extracting meaningful features, while the small set of collected data prevented the use of DL techniques. We thus chose a “hybrid” approach, feeding each spatiotemporal data point (as an independent feature) to SVMs, which are known to display low sensitivity to irrelevant features⁴⁸ and are one of the most widely used state-of-the-art techniques for EEG decoding.^{49–51} While several automatic and semi-automatic denoising techniques that could (at least in principle) improve our decoding performance exist, their contribution is debatable in the general case.⁴²

Another aspect contributing to the choice of a relatively simple ML algorithm is the speed of prediction, which would have been substantially higher with DL or even recurrent architectures. Decreasing the

prediction speed decreases the number of stimuli that can be administered during an optimization run, which could lead to worse optimal stimuli than in the case of longer optimization runs with higher performance decoders. The interplay between decoder performance and prediction time and thus the ability to explore more candidate optima should be tackled in future developments of the present work.

Here, we gave a quantitative idea of what tasks lead to a successful decoding (corners, five vertical and horizontal lines) and which ones are too difficult given our simple decoding strategy (luminosity, out vs center). The two most difficult tasks were characterized by an elevated number of target classes (luminosity) or by the concurrent variation of stimulus size and location. The latter is particularly difficult as, in general, stimuli that occupy more peripheral portions of the visual field excite a lower number of photoreceptor and thus of neurons in higher areas because of the foveated nature of the human retina,^{52,53} which also happens in the case of smaller visual stimuli. Thus, we have a partial compensation of the two effects of moving peripherally and increasing the size of a stimulus, negatively affecting the decoder performance.

When we referred to the decoding of the stimulus size, what was actually decoded depended on a combination of size, luminosity, and luminance. We could have chosen stimuli where size was the only feature that varied, e.g., by using a black-and-white checkerboard pattern on a gray background, as in most studies in the neuroscience of vision field. Here, we did not apply a correction for stimulus luminosity as we were not interested in determining the neural correlates of given visual features in isolation, such as stimulus size and location. Instead, we strived to determine whether EEG could be a technique capable of discriminating visual stimuli compatible with ES and simple enough to be composed to build and study simplified visual scenes in the future. ES produces solid light blobs whose luminosity, size, and location are intertwined and related to the chosen stimulating contact in the implanted electrode and the stimulation protocol and amplitude.¹⁶ Crafting a set of visual stimuli in which single stimulus features are maximally decoupled and distinguishable could still be of translational interest as it could allow the implementation of better feature extractors preceding decoding. Other than increasing the resulting decoding accuracy, this could reduce the number of trials required to attain a given performance, speeding up the process of stimulation optimization and thus decreasing the burden for the target subject.

Multi-trial decoding generally led to an increase in performance, except in cases in which the single-trial performance was not significantly higher than chance. In those cases, we did not observe systematic worsening of the performances but rather unpredictable behavior. All of this is reasonable, since the presented multi-trial approach employs calibrated classification probabilities. When the probability of correct classification is not above chance, the classification probabilities are not informative and the outcome when merging is unpredictable. Moreover, the above observation serves as a rule of thumb for establishing when to employ multi-trial decoding: when single-trial performance is significantly above chance.

While the visual stimuli employed in this work are much simpler than the visual scenes that the healthy is confronted with in everyday life, they are not so far from what is possible through current visual neuroprostheses.^{6,54,55} In particular, existing visual neuroprostheses provide limited resolution, given by inter-electrode crosstalk and by the hardware complexity to accommodate high electrode counts in the

retina or optic nerve because of anatomical reasons.⁵⁶ To cope with such limited resolution, it is customary to heavily simplify the scenes coming from an external camera, focusing on reconstruction of the shapes and locations of a few salient objects and omitting most of the structural details.^{57,58}

We remark that decoders trained on a cohort of healthy subjects may not be usable out of the box on patients because of the differences between the cortical correlates of electrical and visual stimulation and of the possible plastic modifications that occurred in patients after the insurgence of blindness.^{59,60} In both cases, extensive characterization of the neuroprosthetic device after implantation will be needed, mapping simple stimuli to the evoked sensations determined through patient feedback. Then, transfer learning in the form of fine-tuning of the decoders on patient data (correspondence between visual perception evoked through electrical stimulation and corresponding evoked EEG) and domain adaptation techniques (matching the distribution of responses from healthy subjects and target patient) could be used to facilitate generalization.

METHODS

Experimental setup

EEG signals were acquired from a 128-channel gel-based ANT Neuro cap using eego64TM software. We acquired data from ten healthy subjects (3 females), 23 ± 2 years old (average \pm standard deviation). All participants had normal or corrected-to-normal vision and no history of neurological disorders. They were positioned in front of a 75 in. screen ($166 \times 93 \text{ cm}^2$) at a distance of 40 cm, so that the screen covered a field of view of 100° horizontally and 128° vertically [Fig. 1(b)].

Each subject performed two experimental sessions lasting approximately one hour on different days. Each experimental session was divided into 30 sub-sessions consisting of the administration of 50 visual stimuli each. Subsequent sub-sessions were separated by a resting period controlled by the subject (on the order of a few minutes). Each visual stimulus was flashed in and kept for 750 ms, then flashed off. The next stimulus was flashed in after an inter-stimulus time randomly chosen in the range of (1000 and 1250) ms [Fig. 1(c)]. During each sub-session, subjects were asked to fixate a red dot in the center of the screen so that the administered visual stimuli spanned their whole visual field.

For each experimental session, we administered 25 repetitions of each one of 60 different visual stimuli, representing gray (luminosity equals 0.5) squares and rectangles with different sizes, locations, and orientations (vertical and horizontal) on a black background (luminosity equals 0). The complete set of visual stimuli is shown in [supplementary material](#) Fig. 1.

Data preprocessing

For each session, we removed EEG channels with an impedance higher than 100 k Ω (5 channels on average). We acquired signals at a sampling frequency of 1 kHz. We bandpass filtered the continuous recordings from each session between 0.5 and 30 Hz to remove low frequency drifts, including bands up to beta. Specifically, we used the `mne.Raw.filter` function to create a one-pass, zero-phase, non-causal FIR filter designed using window method with Hamming window, with lower and upper cutoff frequencies (defined as half-amplitude cutoff in the middle of the transition band) 0.25 and 33.75 Hz,

respectively. For each stimulus presentation, we considered the signal between -0.3 and 0.75 s, with respect to the stimulus onset. We will refer to the signal between 0 and 0.75 s as an epoch or trial. In symbols,

$$X_j^{nm}[t] = X_j^m[\lfloor (T_n - 0.3) \cdot f_s \rfloor + t],$$

$$\text{for } t = 1, \dots, \lfloor 1.05 \cdot f_s \rfloor \in \mathbb{N},$$

where $\lfloor \cdot \rfloor$ indicates the floor function, thus $\lfloor \tau \cdot f_s \rfloor$ refers to the sample number at time τ seconds. T_n is the time of occurrence of the n th stimulus presentation, while j indicates the recording channel, and m the subject.

We performed baseline correction by subtracting from each epoch the average value of its baseline, taken in the interval $(-0.3, -0.1)$ s,

$$X_j^{nm}[t] \leftarrow X_j^{nm}[t + \lfloor 0.3 \cdot f_s \rfloor] - \frac{1}{\lfloor 0.2 \cdot f_s \rfloor} \sum_{i=1}^{\lfloor 0.2 \cdot f_s \rfloor} X_j^{nm}[i].$$

Then, we performed another step to reject bad channels: for each channel, we computed the average of the RMS value over all epochs, or

$$aRMS_j^m = \frac{1}{N_n} \sum_{n=1}^{N_n} \left\{ \sqrt{\frac{1}{N_t} \sum_{t=1}^{N_t} (X_j^{nm}[t])^2} \right\}.$$

We removed from the dataset all channels with outlier average RMS (aRMS), namely, all the values of j so that

$$aRMS_j^m > q\left(\left\{aRMS_j^{m,j'}\right\}, 3\right) + 1.5 \cdot \text{IQR}\left(\left\{aRMS_j^{m,j'}\right\}\right), \quad \text{or}$$

$$aRMS_j^m < q\left(\left\{aRMS_j^{m,j'}\right\}, 1\right) - 1.5 \cdot \text{IQR}\left(\left\{aRMS_j^{m,j'}\right\}\right),$$

where $q(S, k)$ indicates the k th quartile of the set of values S , and IQR is the interquartile range,

$$\text{IQR}(S) = q(S, 3) - q(S, 1).$$

Stimulus classes

In order to assess the possibility to discriminate several features of a stimulus, we defined a set of decoding tasks, respectively, to decode information about the location of the presented stimulus, its size, or both. To do this, we formed classes from different sets of the 60 original stimuli, and we determined ten decoding tasks where it was necessary to discriminate whether a stimulus belonged to one of several mutually exclusive classes. The set of stimulus classes and decoding tasks is detailed in [supplementary material](#) Fig. 2. The decoding tasks “left vs right” and “superior vs inferior” discriminated stimuli belonging, respectively, to the left or right portion and superior or inferior portion of the screen grid and thus to the subject visual field. The tasks “vertical bars” and “horizontal bars” insert a finer granularity in the decoding of the vertical or horizontal location of a stimulus in the subject visual field, introducing five levels (two blocks left to the center, one block left to the center, center, one block right to the center, two blocks right to the center). The tasks “corners” and “outer vs middle” allow to further investigate the possibility of decoding different locations in the visual field. “Bars,” “blocks,” and “luminosities” allow to decode stimulus sizes on, respectively, 3, 3, and 5 levels. “Outer vs

center” is used to study the decoupling of stimulus size and location, trying to distinguish large stimuli in the outer part of the visual field and small stimuli in the central part of the visual field.

Event related potential (ERP) analysis

ERPs were computed by averaging multiple epochs belonging to the same class, i.e.,

$$Y_j^{C,m}[t] = \frac{1}{N_C} \sum_{n \in C} X_j^{n,m}[t],$$

where C indicates the class, m the subject, j the channel, t the time step, and N_C the number of samples from class C . A statistical analysis was performed using the MNE function `mne.stats.permutation_cluster_test` with default parameters to see if there was a significant difference between the responses from two or multiple classes. This analysis was done separately for each subject and recording channel in the occipital and parietal lobes (identified as the channels containing an “O” or a “P” in their identifier).

We chose to show the results relative to subjects 1, 2, and 10 as they yielded the best, worst, and average single-trial decoding accuracies, respectively.

Single- and multi-trial decoding

Classification was done using a support vector classifier (SVC)⁶¹ calibrated using the function `sklearn.calibration.CalibratedClassifierCV` with default parameters (where the calibration happens on the training set).

The spatiotemporal matrix of EEG signals for a trial was reshaped into a one-dimensional array of features, concatenating the time-signal from each EEG channel, considering all channels (not just the occipital and parietal ones). Thus, the EEG signal values corresponding to each couple of time points and channels were treated as an independent feature in the classification.

When the classes used in a decoding task were composed of approximately the same number of stimuli, i.e., the corresponding dataset was balanced, we estimated the decoding accuracy for the given decoding task using a tenfold cross-validation. Balanced decoding tasks are ideally composed of the same number of trials per class, even though on average four trials out of the 3000 acquired for each subject were not recorded by the acquisition software, thus resulting in a possible mismatch of a few trials across different classes. Instead, when the dataset corresponding to a given decoding task was heavily unbalanced, the accuracy was estimated using cross validation employing 10 random resamplings (without replacement) of the trials from each class until the number of trials of the least represented class for each decoding task was reached. The resulting number of samples for each decoding task is detailed in [supplementary material Table 1](#).

We standardized the signals in the training and test sets using z-scoring as follows:

$$X_{i,j}^{n,m} \leftarrow \frac{X_{i,j}^{n,m} - \mu_j^m}{\sigma_j^m},$$

where we computed the mean and standard deviation for each subject (m) and channel (j) on the training set (for standardizing both training and test set) as

$$\mu_j^m = \frac{1}{N_i \cdot N_n} \sum_t \sum_n X_j^{n,m}[t],$$

$$\sigma_j^m = \sqrt{\frac{1}{N_i \cdot N_n - 1} \sum_t \sum_n (X_j^{n,m}[t] - \mu_j^m)^2},$$

where N_i is the number of time steps and N_n is the number of trials.

In order to determine whether our decoding accuracy was significantly above chance, we compared the decoding accuracy obtained for each task and subject to the accuracy obtained by a random decoder. Such a decoder produced the same number of decoding labels as the given subject for the selected decoding task. We thus compared the ten accuracy values corresponding to the ten different folds/resamplings both for the real subject and for the random decoder using a Wilcoxon test, after checking that each set of ten accuracy values was not Gaussian. The obtained p-values were corrected using Bonferroni correction for the number of subjects, considering each task as an independent scenario.

Additionally, we computed a sensitivity map of our RBF SVMs to the input features, providing a measure of feature importance, using the method introduced in Ref. 62 and employed for EEG decoding in Ref. 63. The computation of such maps was performed using the code from Ref. 63 freely accessible at <https://github.com/gretatuckute/DecodingSensitivityMapping/>.

In order to compute the predicted labels through multi-trial decoding, for each fold, we considered the test set and predicted the classification probabilities for each sample therein. Then, we combined such probabilities using a sliding window of N trials on the shuffled test set, with N ranging from 2 to 15. Accuracies were calculated based on the combined probabilities using an `argmax` function, as for single-trial.

The aggregated probability, p_k , to belong to macro-class k is obtained from the equation

$$p_k = \frac{\prod_{i=1}^N \left(\frac{p_{i,k}}{p_{i,M}} \right)^{1/N}}{\sum_{j=1}^M \left[\prod_{i=1}^N \left(\frac{p_{i,j}}{p_{i,M}} \right)^{1/N} \right]}, \quad k = 1, \dots, M.$$

Here, i is the trial and N the total number of trials to aggregate (here, ranging from 1 to 15). M is the number of alternative macro-classes constituting the given tuning set (e. g., $M = 2$ for coarse left vs right and $M = 5$ for fine left vs right).⁴¹ The original probabilities, $p_{i,j}$, were obtained after calibrating the trained SVMs through Platt’s scaling.⁶⁴ The method used to combine probabilities did not affect calibration.⁵⁵

Independent component analysis

Even though subjects were instructed to fixate the center of the screen during the whole experiments, residual movement activity might influence decoding scores. To test and ensure that the decoding was based only on the neural correlates of the stimuli, rather than eye movements, we also calculated, for subjects 1, 2, and 10, the decoding accuracy after removing ocular activity (blinks and horizontal eye movements) by means of independent component analysis (ICA).⁶⁵ As ICA algorithm we used the Picard ICA algorithm within the MNE toolbox for Python (function `mne.preprocessing.ICA`).

SUPPLEMENTARY MATERIAL

See the [supplementary material](#) for further information on experimental design, effect of eye artifact removal on decoding performance, feature importance analyses, and further information on multi-trial performance.

ACKNOWLEDGMENTS

This work was partly funded by the Bertarelli Foundation.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Ethics Approval

Ethics approval for experiments reported in the submitted manuscript on animal or human subjects was granted. The study was approved by the Commission Cantonale d'Éthique de la Recherche Genève, reference CCER_Etude 2021-01724.

Author Contributions

Simone Romeni and Laura Toni contributed equally to this work.

Simone Romeni: Conceptualization (equal); Formal analysis (equal); Methodology (equal); Software (equal); Writing – original draft (equal). **Laura Toni:** Formal analysis (equal); Methodology (equal); Software (equal); Visualization (equal); Writing – original draft (equal). **Fiorenzo Artoni:** Conceptualization (equal); Methodology (equal); Supervision (equal); Writing – original draft (equal). **Silvestro Micera:** Conceptualization (equal); Funding acquisition (equal); Supervision (equal); Writing – original draft (equal).

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

- M. J. Burton *et al.*, “The Lancet Global Health Commission on Global Eye Health: Vision beyond 2020,” *Lancet Global Health* **9**(4), e489–e551 (2021).
- J. D. Steinmetz *et al.*, “Causes of blindness and vision impairment in 2020 and trends over 30 years, and prevalence of avoidable blindness in relation to VISION 2020: The Right to Sight: An analysis for the Global Burden of Disease Study,” *Lancet Global Health* **9**(2), e144–e160 (2021).
- L. da Cruz *et al.*, “The Argus II epiretinal prosthesis system allows letter and word reading and long-term function in patients with profound vision loss,” *Br. J. Ophthalmol.* **97**(5), 632–636 (2013).
- L. Yue, J. D. Weiland, B. Roska, and M. S. Humayun, “Retinal stimulation strategies to restore vision: Fundamentals and systems,” *Prog. Retinal Eye Res.* **53**, 21–47 (2016).
- C. Veraart *et al.*, “Visual sensations produced by optic nerve stimulation using an implanted self-sizing spiral cuff electrode,” *Brain Res.* **813**(1), 181–186 (1998).
- M. E. Brelén, F. Duret, B. Gérard, J. Delbeke, and C. Veraart, “Creating a meaningful visual perception in blind volunteers by optic nerve stimulation,” *J. Neural Eng.* **2**(1), S22–S28 (2005).
- E. Fernández *et al.*, “Visual percepts evoked with an intracortical 96-channel microelectrode array inserted in human occipital cortex,” *J. Clin. Invest.* **131**(23), e151331 (2021).
- J. de Ruyter van Steveninck *et al.*, “Real-world indoor mobility with simulated prosthetic vision: The benefits and feasibility of contour-based scene simplification at different phosphene resolutions,” *J. Vision* **22**(2), 1–14 (2022).
- N. Han, S. Srivastava, A. Xu, D. Klein, and M. Beyeler, “Deep learning-based scene simplification for bionic vision,” in *Proceedings of the Augmented Humans International Conference (AHs '21)* (Association for Computing Machinery, 2021), pp. 45–54.
- S. Yang, D. Han, J. Wu, X. Chen, and V. Chung, “Scene simplification for simulated prosthetic vision with improved scene understanding,” in *IEEE International Conference on Consumer Electronics (ICCE)*, 2024.
- V. Vergnieux, M. J.-M. Macé, and C. Jouffrais, “Simplification of visual rendering in simulated prosthetic vision facilitates navigation,” *Artif. Organs* **41**(9), 852–861 (2017).
- J. Wang *et al.*, “The application of computer vision to visual prosthesis,” *Artif. Organs* **45**(10), 1141–1154 (2021).
- S. C. Chen, G. J. Suaning, J. W. Morley, and N. H. Lovell, “Simulating prosthetic vision: I. Visual models of phosphenes,” *Vision Res.* **49**(12), 1493–1506 (2009).
- S. C. Chen, G. J. Suaning, J. W. Morley, and N. H. Lovell, “Simulating prosthetic vision: II. Measuring functional capacity,” *Vision Res.* **49**(19), 2329–2343 (2009).
- M. Beyeler, D. Nanduri, J. D. Weiland, A. Rokem, G. M. Boynton, and I. Fine, “A model of ganglion axon pathways accounts for percepts elicited by retinal implants,” *Sci. Rep.* **9**(1), 9199 (2019).
- J. Delbeke, M. Oozeer, and C. Veraart, “Position, size and luminosity of phosphenes generated by direct optic nerve stimulation,” *Vision Res.* **43**(9), 1091–1102 (2003).
- S. Romeni, D. Zoccolan, and S. Micera, “A machine learning framework to optimize optic nerve electrical stimulation for vision restoration,” *Patterns* **2**(7), 100286 (2021).
- D. L. K. Yamins, H. Hong, C. F. Cadieu, E. A. Solomon, D. Seibert, and J. J. DiCarlo, “Performance-optimized hierarchical models predict neural responses in higher visual cortex,” *Proc. Natl. Acad. Sci. U. S. A.* **111**(23), 8619–8624 (2014).
- M. Schrimpf *et al.*, “Brain-score: Which artificial neural network for object recognition is most brain-like?,” *Neuroscience* (2018).
- S. Tafazoli, C. J. MacDowell, Z. Che, K. C. Letai, C. R. Steinhardt, and T. J. Buschman, “Learning to control the brain through adaptive closed-loop patterned stimulation,” *J. Neural Eng.* **17**(5), 056007 (2020).
- Y. N. Billeh *et al.*, “Systematic integration of structural and functional data into multi-scale models of mouse primary visual cortex,” *Neuron* **106**(3), 388–403 (2020).
- H. C. Stronks, M. P. Barry, and G. Dagnelie, “Electrically elicited visual evoked potentials in Argus II retinal implant wearers,” *Invest. Ophthalmol. Vis. Sci.* **54**(6), 3891 (2013).
- Y. He *et al.*, “Development of an electroencephalogram (EEG) protocol to map electrically-elicited visual responses in blind patients implanted with the Argus II retinal prosthesis,” *Invest. Ophthalmol. Vis. Sci.* **59**(9), 4572 (2018).
- P. Vagni *et al.*, “POLYRETINA restores light responses in vivo in blind Göttingen minipigs,” *Nat. Commun.* **13**(1), 3678 (2022).
- V. Gaillat *et al.*, “Spatially selective activation of the visual cortex via intraneural stimulation of the optic nerve,” *Nat. Biomed. Eng.* **4**(2), 181–194 (2020).
- D. M. Goldenholz *et al.*, “Mapping the signal-to-noise-ratios of cortical sources in magnetoencephalography and electroencephalography,” *Human Brain Mapping* **30**(4), 1077–1086 (2009).
- S. Romeni, G. Marino, L. Pierantoni, and S. Micera, “Decoupling the effects of varying amplitude and frequency in the electrical stimulation of optic nerve fibers,” *Curr. Dir. Biomed. Eng.* **9**(1), 194–197 (2023).
- S. Bagchi and D. R. Bathula, “EEG-ConvTransformer for single-trial EEG-based visual stimulus classification,” *Pattern Recognit.* **129**, 108757 (2022).
- A. X. Stewart, A. Nuthmann, and G. Sanguinetti, “Single-trial classification of EEG in a visual object task using ICA and machine learning,” *J. Neurosci. Methods* **228**, 1–14 (2014).
- A. T. Gifford, K. Dwivedi, G. Roig, and R. M. Cichy, “A large and rich EEG dataset for modeling human visual object recognition,” *NeuroImage* **264**, 119754 (2022).
- M. Taghizadeh-Sarabi, M. R. Daliri, and K. S. Niksirat, “Decoding objects of basic categories from electroencephalographic signals using wavelet transform and support vector machines,” *Brain Topogr.* **28**(1), 33–46 (2015).

- ³²M. J. Wolff, J. Ding, N. E. Myers, and M. G. Stokes, "Revealing hidden states in visual working memory using electroencephalography," *Front. Syst. Neurosci.* **9**, 123 (2015).
- ³³N. A. McNair, W. C. Clapp, P. Hamm, T. J. Teyler, M. C. Corballis, and I. J. Kirk, "Spatial frequency-specific potentiation of human visual-evoked potentials," *NeuroReport* **17**, 739–741 (2006).
- ³⁴H. Karimi-Rouzbahani, "Evidence for multiscale multiplexed representation of visual features in EEG," *Neural Comput.* **36**(3), 412–436 (2024).
- ³⁵D. V. P. S. Murty, V. Shirhatti, P. Ravishankar, and S. Ray, "Large visual stimuli induce two distinct gamma oscillations in primate visual cortex," *J. Neurosci.* **38**(11), 2730–2744 (2018).
- ³⁶D. V. P. S. Murty *et al.*, "Gamma oscillations weaken with age in healthy elderly in human EEG," *NeuroImage* **215**, 116826 (2020).
- ³⁷Y. Shigihara, H. Hoshi, and S. Zeki, "Early visual cortical responses produced by checkerboard pattern stimulation," *NeuroImage* **134**, 532–539 (2016).
- ³⁸C. Simar *et al.*, "Riemannian classification of single-trial surface EEG and sources during checkerboard and navigational images in humans," *PLoS One* **17**(1), e0262417 (2022).
- ³⁹N. R. Waytowich, Y. Yamani, and D. J. Krusienski, "Optimization of checkerboard spatial frequencies for steady-state visual evoked potential brain-computer interfaces," *IEEE Trans. Neural Syst. Rehabil. Eng.* **25**(6), 557–565 (2017).
- ⁴⁰D. Yang, Y. Liu, Z. Zhou, Y. Yu, and X. Liang, "Decoding visual motions from EEG using attention-based RNN," *Appl. Sci.* **10**(16), 5662 (2020).
- ⁴¹V. A. Satopää, J. Baron, D. P. Foster, B. A. Mellers, P. E. Tetlock, and L. H. Ungar, "Combining multiple probability predictions using a simple logit model," *Int. J. Forecasting* **30**(2), 344–356 (2014).
- ⁴²A. Delorme, "EEG is better left alone," *Sci. Rep.* **13**(1), 2372 (2023).
- ⁴³F. Di Russo, A. Martínez, M. I. Sereno, S. Pitzalis, and S. A. Hillyard, "Cortical sources of the early components of the visual evoked potential," *Human Brain Mapping* **15**(2), 95–111 (2002).
- ⁴⁴H. E. Jones, K. L. Grieve, W. Wang, and A. M. Sillito, "Surround suppression in primate V1," *J. Neurophysiol.* **86**(4), 2011–2028 (2001).
- ⁴⁵S. Kastner, P. De Weerd, M. A. Pinsk, M. I. Elizondo, R. Desimone, and L. G. Ungerleider, "Modulation of sensory suppression: implications for receptive field sizes in the human visual cortex," *J. Neurophysiol.* **86**(3), 1398–1411 (2001).
- ⁴⁶D. Xing, R. M. Shapley, M. J. Hawken, and D. L. Ringach, "Effect of stimulus size on the dynamics of orientation selectivity in Macaque V1," *J. Neurophysiol.* **94**(1), 799–812 (2005).
- ⁴⁷M. E. Stefanou *et al.*, "Late attentional processes potentially compensate for early perceptual multisensory integration deficits in children with autism: Evidence from evoked potentials," *Sci. Rep.* **10**(1), 16157 (2020).
- ⁴⁸C. Bishop, *Pattern Recognition and Machine Learning* (Springer, New York, 2023).
- ⁴⁹M. N. A. H. Sha'abani, N. Fuad, N. Jamal, and M. F. Ismail, "kNN and SVM classification for EEG: A review," in *InECCE2019*, edited by A. N. Kasruddin Nasir, M. A. Ahmad, M. S. Najib, Y. Abdul Wahab, N. A. Othman, N. M. Abd Ghani, A. Irawan, S. Khatun, R. M. T. Raja Ismail, M. M. Saari, M. R. Daud, and A. A. Mohd Faudzi (Springer Singapore, Singapore, 2020), vol. 632, pp. 555–565.
- ⁵⁰I. Guler and E. D. Ubeyli, "Multiclass support vector machines for EEG-signals classification," *IEEE Trans. Inf. Technol. Biomed.* **11**(2), 117–126 (2007).
- ⁵¹M.-P. Hosseini, A. Hosseini, and K. Ahi, "A review on machine learning for EEG signal processing in bioengineering," *IEEE Rev. Biomed. Eng.* **14**, 204–218 (2021).
- ⁵²C. A. Curcio and K. A. Allen, "Topography of ganglion cells in human retina," *J. Comp. Neurol.* **300**(1), 5–25 (1990).
- ⁵³A. B. Watson, "A formula for human retinal ganglion cell receptive field density as a function of visual field location," *J. Vision* **14**(7), 15 (2014).
- ⁵⁴S. Micera *et al.*, "A computational model to design wide field-of-view optic nerve neuroprostheses," *Research Square* (2023).
- ⁵⁵M. Markowitz *et al.*, "Rehabilitation of lost functional vision with the Argus II retinal prosthesis," *Can. J. Ophthalmol.* **53**(1), 14–22 (2018).
- ⁵⁶E. Borda and D. Ghezzi, "Advances in visual prostheses: Engineering and biological challenges," *Prog. Biomed. Eng.* **4**(3), 032003 (2022).
- ⁵⁷J. Wang *et al.*, "Image processing strategies based on a visual saliency model for object recognition under simulated prosthetic vision," *Artif. Organs* **40**(1), 94–100 (2016).
- ⁵⁸H. Li *et al.*, "Image processing strategies based on saliency segmentation for object recognition under simulated prosthetic vision," *Artif. Intell. Med.* **84**, 64–78 (2018).
- ⁵⁹G. E. Legge and S. T. L. Chung, "Low vision and plasticity: implications for rehabilitation," *Annu. Rev. Vision Sci.* **2**(1), 321–343 (2016).
- ⁶⁰N. Sanda *et al.*, "Visual brain plasticity induced by central and peripheral visual field loss," *Brain Struct. Funct.* **223**(7), 3473–3485 (2018).
- ⁶¹C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.* **2**(3), 1–27 (2011).
- ⁶²P. M. Rasmussen, K. H. Madsen, T. E. Lund, and L. K. Hansen, "Visualization of nonlinear kernel models in neuroimaging by sensitivity maps," *NeuroImage* **55**(3), 1120–1131 (2011).
- ⁶³G. Tuckute, S. T. Hansen, N. Pedersen, D. Steenstrup, and L. K. Hansen, "Single-trial decoding of scalp EEG under natural conditions," *Comput. Intell. Neurosci.* **2019**, 1–11.
- ⁶⁴J. Platt, "Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods," in *Advances in Large-Margin Classifier* (MIT Press, 2000), Vol. 10.
- ⁶⁵M. Chaumon, D. V. M. Bishop, and N. A. Busch, "A practical guide to the selection of independent components of the electroencephalogram for artifact correction," *J. Neurosci. Methods* **250**, 47–63 (2015).