Title

Contrast-dependent response modulation in convolutional neural networks captures behavioral and neural signatures of visual adaptation

Abbreviated title

Capturing human visual adaptation with CNNs

Authors

Amber Marijn Brands[1*], Zilan Oz[1], Nikolina Vukšić[1], Paulo Ortiz[1], and Iris Isabelle Anna Groen[1,2]

Affiliations

[1] Informatics Institute, University of Amsterdam, Amsterdam, The Netherlands

[2] Department of Psychology, University of Amsterdam, Amsterdam, The Netherlands

[*] Corresponding author

Number of pages

38

Number of figures

9

Number of words

abstract: 249 words

introduction: 648 words

discussion: 1496 words

The authors have declared that no competing interests exist

Abstract

Human perception remains robust under challenging viewing conditions. Robust perception is thought to be facilitated by nonlinear response properties, including temporal adaptation (reduced responses to repeated stimuli) and contrast gain (shift in the contrast response function with pre-exposure to a stimulus). Temporal adaptation and contrast gain have both been shown to aid object recognition, however, their joint effect on perceptual and neural responses remains unclear. Here, we collected behavioural measurements and electrocorticography (EEG) data while human participants (both sexes) classified objects embedded within temporally repeated noise patterns, whereby object contrast was varied. Our findings reveal an interaction effect, with increased categorization performance as a result of temporal adaptation for higher but not lower contrast stimuli. This increase in behavioral performance after adaptation is associated with more pronounced contrast-dependent modulation of evoked neural responses, as well as better decoding of object information from EEG activity. To elucidate the neural computations underlying these effects, we endowed deep convolutional neural networks (DCNN) with various temporal adaptation mechanisms, including intrinsic suppression and temporal divisive normalisation. We demonstrate that incorporating a biologically-inspired contrast response function to modify temporal adaptation helps DCNNs to accurately capture human behaviour and neural activation profiles. Moreover, we find that networks with multiplicative temporal adaptation mechanisms, such as divisive normalization, show higher robustness against spatial shifts in the inputs compared to DCNNs employing additive mechanisms. Overall, we reveal how interaction effects between nonlinear response properties influence human perception in challenging viewing contexts and investigate potential computations that mediate these effects.

Significance statement

Humans are able to perceive the environment even when viewing conditions are suboptimal. This robust perception has been linked to nonlinear neural processing of incoming visual information. Here, we examine the joint impact of two neural response properties, temporal adaptation and contrast gain, during object recognition, demonstrating interaction effects on categorization performance and in evoked neural responses. Using convolutional neural networks, we investigate various temporal adaptation mechanisms mediating the neural responses and perception, demonstrating that introducing contrast-dependent modulation of the unit activations captures human behaviour and neural object representations. Our findings shed light on how neural response properties give rise to robust perception and offer a framework to study the underlying neural dynamics and their impact on perception.

Introduction

Human perception is highly robust under challenging viewing conditions, illustrated for instance by our ability to detect approaching cars on misty roads (**Fig. 1A**). How does the visual cortex obtain robust representations in such challenging viewing conditions? Earlier studies have identified several neural phenomena that may optimize neural processing and aid in perception.

First, when stimuli are repeated, neural responses to subsequent presentations are reduced (Lueschow et al. 1994; Groen et al. 2022; for a review, see Grill-Spector et al. 2006), with strongest reductions for similar as opposed to different inputs (Sawamura et al. 2006; Brands et al. 2024; **Fig. 1B**). This form of temporal adaptation, known as repetition suppression, has been linked to increased recognition performance (Vinken et al. 2020; for a review, see Henson and Rugg 2003). Second, neural response magnitudes adapt to match the variability in contrast of incoming visual inputs, referred to as contrast gain, evident as a shift of the dynamic range of the contrast response function (CRF) as a result of pre-exposure to an adapter (Wilson and Humanski 1993; Dao et al. 2006; **Fig. 1C**). Contrast gain is thought to enable perceptual constancy despite changes in stimulus contrast (Foley and Legge, 1981; Sclar and Freeman, 1982).

Despite extensive study of these neural response properties in isolation, it remains unclear how they operate in a joint fashion, as is often the case in naturalistic settings. A study by Cravo et al. (2013) showed that temporal adaptation enhances contrast sensitivity of targets via temporal expectation, linking these perceptual effects to delta oscillations in EEG. Here, we investigate if such effects can also be detected in evoked potentials, which show modulation both by repeating inputs (Summerfield et al., 2011; Schweinberger and Neumann, 2016) and contrast (Vassilev et al., 1994; Schadow et al., 2007).

Another outstanding question is what neural computations underlie temporal adaptation. Many nonlinear response properties are thought to be governed by canonical computations, arising from intrinsic biophysical mechanisms operating within individual neurons (Whitmire and Stanley, 2016) or from recurrent interactions (del Mar Quiroga et al., 2016). Previous work by Vinken et al. (2020) implemented temporal adaptation in a deep convolutional neural network (DCNN) via an additive adaptation state into each individual unit, thereby capturing various neurophysiological and behavioral phenomena which are often thought to require recurrent network mechanisms operating in a neural circuit. The authors argue that temporal adaptation on the level of the single unit, as opposed to recurrence, provides a simpler solution, yet is still sufficient to model adaptation effects.

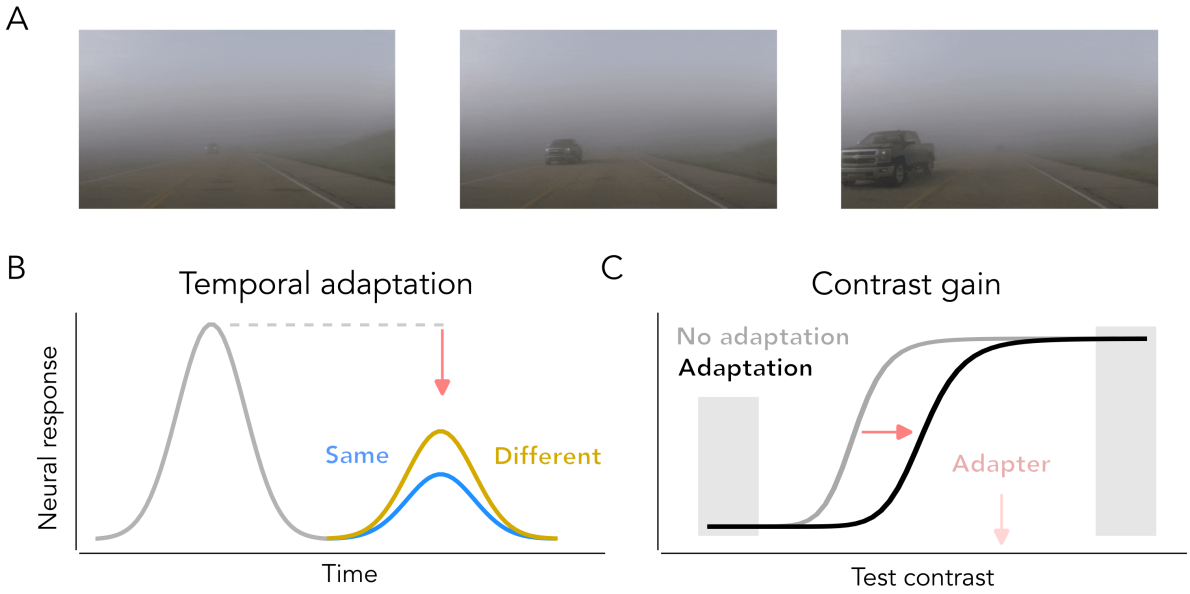

Figure 1: Nonlinear response properties in visual cortex. **A**: Humans are able to recognize targets, for example a car, in challenging viewing conditions, for example mist. **B-C**: Nonlinear neural response properties studied here. Temporal adaptation (B) refers to the reduction of neural responses when stimuli are repeated, whereby response reductions are more pronounced for similar compared to different repeated inputs. Contrast gain (C) refers to the shift of the dynamic range of the contrast response function when presented with an adapter prior to a test stimulus. Responses saturate for lower and higher contrast levels (grey shaded areas).

An alternative, biophysically-realistic form of temporal adaptation that can operate on the single-unit level is divisive normalization (Heeger, 1992, 1993), which divides neural responses by a common factor, serving as a promising candidate for wide range of operations, including both temporal adaptation and contrast gain (for a review, see Carandini and Heeger 2012). Here, we investigate divisive normalization in the context of robust object recognition, and test whether this neural mechanism, when implemented in task-driven DCNNs, offers advantages over other forms of adaptation.

To measure the effects of temporal adaptation and contrast gain on perception, human participants performed an object recognition task with repeated noise patterns, while electrocorticography (EEG) measurements were collected. We find an interaction effect, with improved categorization performance when adapting to same noise at higher contrast levels. This behavioral effect is reflected in evoked responses showing stronger contrast-dependent modulation when adapting to same, as opposed to different or no noise. We demonstrate that DCNNs implemented with temporal adaptation accurately match human behavior and neural response profiles by incorporating a biologically-inspired contrast response function. Moreover, we find that the multiplicative nature of divisive normalization results in higher robustness against spatial shifts in the input compared to DCNNs using additive computations. Together, these

results demonstrate that robust perception during object recognition arises from an interaction between temporal adaptation and contrast gain, and offer an image-computable framework to investigate the neural mechanisms mediating perceptual adaptation.

Materials and Methods

Participants

Twenty-four subjects participated in this study (age mean $\pm$ SD, 21 $\pm$ 2 years, 2 males), which was approved by the Ethical Committee of the Psychology Department at the University of Amsterdam. All participants gave written informed consent before participation and were rewarded with study credits or financial compensation (10 euro/h). Two participants were excluded from analysis because of incomplete recordings, and one participant was excluded because of technical issues that arose during the recording session, yielding a total of 21 included participants.

Stimuli

Participants were instructed to categorize exemplars of four different digits (3, 6, 8 and 9) from the MNIST dataset (LeCun et al. 2010) presented with varying object contrast levels (50%, 60%, 70%, 80% and 90%) (**Fig. 2A**). The digit was embedded in noise, which consisted of adding a pixelized grey-scale noise pattern which was drawn from a uniform distribution on interval $[0, 1)$ following a pink distribution (such that the power spectral density was inversely proportional to the frequency of the signal). The image consisting of the noise-embedded digit will from here on be referred to as the *test* image. Varying the contrast of the digit resulted in easy (high contrast) and difficult (low contrast) task conditions (Root Mean Square test image, mean 50% = 0.45, SD = 0.018; mean 60% = 0.46, SD = 0.018; mean 70% = 0.47, SD = 0.019; mean 80% = 0.48, SD = 0.024; mean 90% = 0.49, SD = 0.026).

Experimental procedure

We adapted the object categorization task used in Vinken et al. (2020), who used hand-drawn doodles hidden in a temporally repeated noise pattern. Here, we extended the paradigm by presenting the targets at various contrast levels and replacing the hand-drawn doodles with MNIST digits. Participants completed one EEG recording session lasting approximately 90 minutes consisting of two blocks with a total of 880 trials. Participants were instructed to fixate on a red cross in the center of a grey screen throughout each trial. Categorization responses were acquired by keyboard presses and inter-trial intervals were sampled from a uniform distribution between 1.25 and 1.75 seconds. After every 50 trials, a break was inserted during which participants were encouraged to take a short rest.

During the first block, participants were presented with three different trial types. Two trial types consisted

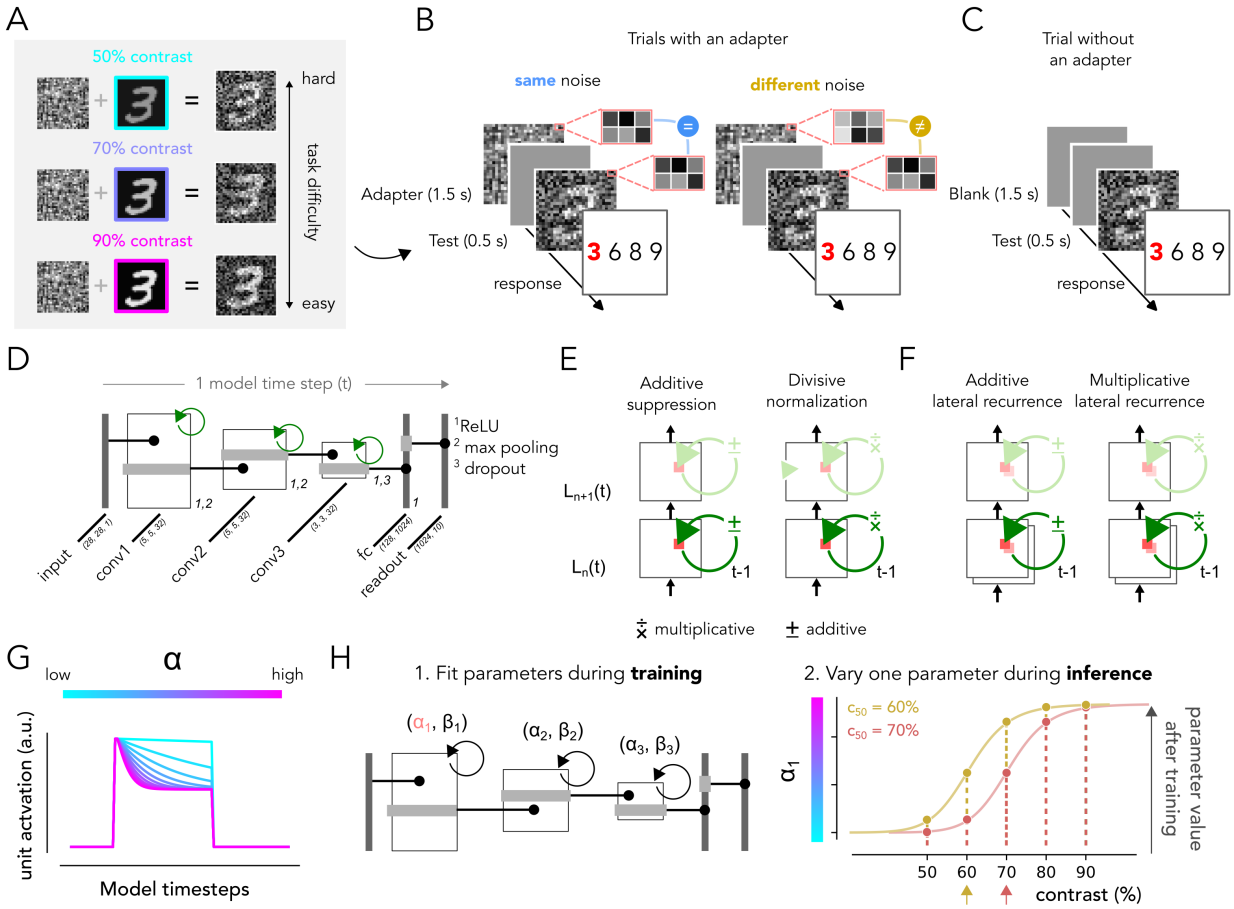

Figure 2: Experimental design and network modeling. A: Object recognition task with the test images consisting of objects (digits 3, 6, 8 and 9 from the MNIST dataset) embedded in a pixelised noise pattern. Contrast of the digit image was varied (50%, 60%, 70%, 80% and 90%). **B**: Adaptation trials, consisting of the presentation of same (left) or different (right) noise prior to the test image. **C**: Control trials, consisting of the presentation of the test image in isolation. **D**: Deep convolutional neural networks (DCNNs) were trained with a feedforward backbone consisting of three convolutional filter layers, one fully connected layer and a readout layer. After each convolution, history-dependent adaptation was applied (depicted in green), feeding activations from the previous model time step, i.e. previous feedforward pass. Filter sizes are represented within parentheses. **E**: Temporal adaptation mechanisms which operate on the level of the single unit. Left, Additive suppression introduced previously by Vinken et al. (2020) whereby each unit feeds its activations from the previous timestep using an additive approach. Right, Divisive normalization introduced by Heeger (1992, 1993) which feeds activations from the previous timestep using a multiplicative approach. **F**: Temporal adaptation mechanism which feed unit activations across feature maps either through additive (left) or multiplicative (right) interactions inspired by Vinken et al. (2020). **G**: Effect of varying a parameter controlling the temporal adaptation (here, α of additive suppression). A higher value of α results in larger contribution of previous inputs and a stronger decay of the response. **H**: Introducing contrast-dependent modulation of unit activations. *Left*, During training all parameters controlling the temporal adaptation (e.g., α and β for additive suppression) are fitted for every layer on full contrast trials (never seen by participants during the EEG experiment). *Right*, During inference, temporal parameters of interest (e.g. α) in the first convolutional layer are adjusted based on a contrast response function described by Eq. 10. The maximal response (grey arrow) is set to the parameter value after training. c_{50} denotes the position of the CRF along the contrast axis (arrow), which is varied during inference leading to different adjustments of the temporal parameter values.

of the presentation of an adapter (1.5 s), followed by the test image (0.5 s), with the adapter containing the same noise as that of the subsequent test image or a randomly generated different noise pattern (**Fig. 2B**). We refer to these trials where same or different noise was presented prior to the test image as *same-noise* and *different-noise* trials, respectively. The time in between the adapter and test was fixed (134 ms) and for each trial, a different exemplar was selected and a unique noise pattern was created, such that there were no repetitions across trials. The third trial type served as a control - referred to as the *blank* trial - which consisted of a blank image followed by the presentation of the test image (**Fig. 2C**). For the *same-noise*, *different-noise* and *blank* trials eight exemplars per digit were presented for each contrast condition, resulting in a total of 480 trials (160 per adapter type, i.e. four targets $\times$ five contrast levels $\times$ eight exemplars).

In order to conduct a decoding analysis (see below) a second block of trials was presented at the end of the session during which participants categorized the target digits (100% contrast) without added noise. Here, 100 exemplars were presented for each digit class, resulting in a total of 400 trials.

EEG acquisition and preprocessing

EEG recordings were made with a Biosemi 64-channel Active TwoEEG system (Biosemi Instrumentation; www.biosemi.com), using an extended 10–20 layout modified with two additional occipital electrodes (I1 and I2, while removing electrodes F5 and F6). Eye movements were monitored with electro-oculograms (EOGs). The EEG signal was digitized at 1024 Hz sampling rate. Recording was followed by offline referencing to external electrodes placed on the mastoids. Preprocessing for the purpose of computing event-related responses (EPRs) was done in Python, MNE (Gramfort et al., 2013) and included the following steps used in standard pipelines in our lab (e.g., (Groen et al. 2012, see preprocessing_createEpochs.ipynb): a high-pass filter at 0.01 Hz (6 dB/octave); a low-pass filter at 30 Hz (6 dB/octave) (Butterworth bandpass zero-phase, two-pass forward and reverse, non-causal filter; filter order 16, effective, after forward-backward; cutoffs at 0.01, 30.00 Hz: -6.02, -6.02 dB); a notch filter (zero-phase) at 50 Hz; epoch segmentation in -100 to 500 ms from stimulus onset of the test image; baseline correction between -100 and 0 ms; ocular correction using the EOG electrodes (Gratton et al., 1983); conversion to Current Source Density responses (Perrin et al., 1987).

After computing 880 trials per subject we performed epoch selection for individual electrodes and subjects (preprocessing_epochSelection.py): We first computed the maximum response in the [-0.1, 0.5] time-window (voltage time courses were first full-wave rectified), after which the standard deviation (SD) of

these maximum values over all trials was computed. Trials were excluded from analysis if the maximum response was > 2 SD. Based on the selection methods described above, on average 6.7% (min: 4.4%, max: 8.2%) of the epochs were excluded.

Data analysis

Statistical testing for behavioral data Performance (quantified as the categorization accuracy) and reaction times were computed separately for each subject and trial-type, and subsequently fitted with linear mixed effects models (LMMs) using the *lme4* package implemented in R (Bates and DebRoy, 2004) (statistics_lmer.R). We choose an LLM approach because humans are known to exhibit individual differences in contrast sensitivity (Baker and Graf, 2009), resulting in substantial variation in categorization performance across subjects. The model comprised two fixed effects: adapter type (none, same or different) and contrast level (50%, 60%, 70%, 80% and 90%), and one random effect (subject). The estimated coefficients for each of the fixed effects were evaluated with ANOVAs, and the resulting p-values were corrected for multiple comparisons across the three adapter types using Tukey method for comparing three estimates (i.e. the adapter types).

Averaging procedure for ERP analysis In total, we obtained 880 single ERPs resulting from presentation of the test image, including 480 obtained from the noise-embedded object trials and 400 obtained from clean object trials. For the ERP analysis responses were averaged over trials as follows: For the effect of adapter type, trials were averaged over exemplars, digit classes and contrast level, yielding 160 ERPs per adaptation condition. For the effect of contrast level, trials were averaged over exemplars, digit classes and adapter types, yielding 96 ERPs per contrast condition. For the interaction effect between adapter type and contrast, trials were averaged over exemplars and digit classes yielding 32 ERPs per condition.

Identification of electrodes of interest We focus our analyses on electrodes recording from early visual areas, including Oz, Iz, O1 and O2, and from higher-order visual areas, including P9 and P10, based on prior literature showing object-related processing/repetition suppression effects at these electrodes (Gruber and Müller, 2002; Schweinberger and Neumann, 2016; Johnsdorf et al., 2023). Moreover, for examining the contrast-dependent modulation, we also inspected parietal-occipital electrodes, including Pz, P1, P2, P3, and P4, which have been implicated in gain control contrast mechanisms (Di Russo et al., 2001). For the decoding analysis (see Modeling section below), single ERPs of all electrodes were used, motivated by the fact that this required no *a priori* assumptions regarding the spatial distribution of object-related representations.

Identification of time windows of interest To study the effect of adapter types, we first determined which time windows are affected by preceding inputs. We achieved this by subtracting the ERPs of trials without adaptation from the trials with adaptation separately for same- and different-noise conditions. Subsequently, we determined the timepoints for which the difference was significantly different from 0 using a one-sample T-test. To investigate the effect of contrast, we selected time windows implicated in early and late visual processing. The time window implicated in early visual processing including the P100, with an onset of 100-150 ms after stimulus onset, a time window which has been linked to core object recognition whereby a first wave of image-selective neural activity is present throughout higher-order areas (Thorpe et al., 1996; DiCarlo and Maunsell, 2000; Philiastides and Sajda, 2006). To study late visual processing, we focused on the P300 component (computed time window of 280-320 ms), which is known to reflect the engagement of higher-order cognitive functions, including decision-making behavior during visual processing (Nazari et al., 2010; Philiastides and Sajda, 2006).

ERP decoding analysis To examine how adaptation and contrast gain jointly affect representations of object classes in noise in EEG responses, we ran a series of decoding analyses, which consisted of training a linear classifier to predict the presented digit (i.e. 3, 6, 8 or 9) based on the spatially-distributed ERP responses. We performed a subject-wise decoding analysis separately for the ERPs obtained by presenting either clean or noise-embedded test images (mkFigure8AB.ipynb). We used the Python sklearn function $\mathrm{sklearn.svm.SVC()}$, with default parameters, which is able to handle binary as well as multi-class classification. To reduce computation time, ERP signals were first downsampled to 128 Hz. To increase the strength of the signal compared to the background noise a sliding window of 62.5 ms was used. For the trials presenting the noise-embedded test images (480 in total), a k-fold cross-validation was applied with 15 folds during which each ERP was used for inference once, with 448 images used for training and 32 images used for testing per fold. Here, the images used for testing within each fold belonged to one contrast level and one adapter type (e.g. same-noise adapter for an object contrast of 50%). For clean test images, the classifier was trained on 360 trials and tested on 40 trials. Similarly to the trials presenting the noise-embedded objects, a k-fold cross-validation was performed with 10 folds such that each trial was used once during inference.

DCNN modeling

Deep neural network implementation

DCNN models were trained on a same object recognition task as the participants with either same or different noise preceding the test image. All networks were implemented using Pytorch (Paszke et al., 2019) with standard convolutions and linear layers. We introduced temporal dynamics into the networks by feeding activity from previous timesteps, with one timestep, t, defined as one feedforward sweep. A training sample consisted of three timesteps, during which the networks were fed an adapter image (t_1), followed by a gray-scale image (t_2, pixel values of 0.5) and a test image (t_3), with input size of 1 $\times$ 28 $\times$ 28 (channels $\times$ height $\times$ width) for all timesteps. All pixel values were clamped to the range [0, 1] to reduce numerical complexity and increase computational efficiency. Activations from the last time step were used for the classification, computation of the loss and weight updates. The backbone of all DCNNs consisted of three convolutional layers, one fully connected and a readout layer (**Fig. 2D**, for the implementation see cnn_feedforward.py). We define a linear response $\mathbf{L}_n(t)$ at time t as the output of a convolutional layer n:

$$\mathbf{L}_n(t) = \mathbf{w}_n * \mathbf{x}_{n-1}(t) + \mathbf{b}_n \tag{1}$$

given the unit's current input $\mathbf{x}_{n-1}(t)$, bottom-up convolutional weights $\mathbf{W}_n$ and biases $\mathbf{b}_n$, whereby the convolution operation is represented as $*$. We trained five different model architectures. For the baseline model, no adaptation mechanism was added such that all images were processed independently over time. We then implemented and compared four different adaptation mechanisms (see below), which were applied only in the three convolutional layers. Corresponding parameters were implemented per layer and were optimized simultaneously with regular DCNN parameters. Networks were trained using noise-embedded full-contrast digits and then tested on the same contrast levels which were used for testing the participants (i.e. 50%, 60%, 70% , 80% and 90%). The reason for varying object contrast only during inference was to test networks' ability to generalize to other contrast levels, thereby examining the contrast sensitivity of the networks. This was motivated by the fact that previous research has shown networks are affected by varying this stimulus property (Reith and Wandell, 2020), whereby our aim was to investigate to what degree this sensitivity results in similar categorization performance as is observed in human participants. Networks were trained for 10 epochs (with 60000 images per epoch). To asses the robustness of our findings (Mehrer et al., 2020), 10 instances of DCNNs were trained from random initializations for each training length, resulting in a total of 50 trained models (five model architectures x 10 initializations). Models were trained using the Adam optimizer, a learning rate of 0.001, the cross entropy using stochastic gradient descent with a batch

size of 100. After the last convolution, 50% dropout was applied. Script used for model training can be found at model_train.py.

Additive suppression We implemented an intrinsic suppression mechanism that emerges from intrinsic properties of individual units (**Fig. 2E**, *left*), as introduced previously by Vinken et al. (2020). Here, each unit i in the network has an exponentially decaying intrinsic adaptation state, s_t, which is updated at each time step t, based on its previous state $s(t-1)$ and the previous response $r(t-1)$ as follows:

$$s_i(t) = (1-\alpha)s_i(t-1) + \alpha r_i(t-1) \tag{2}$$

where α is a constant determining the time scale of the decay. This intrinsic adaptation state is then subtracted from the linear response before applying the activation function ϕ.

$$r_i(t) = \phi\left(L_i(t) - \beta s_i(t)\right) \tag{3}$$

where β is a constant scaling the amount of suppression. For $\beta > 0$, these updating rules result in an exponentially decaying response for constant input that recovers in case of no input.

Divisive normalization Divisive normalization was originally proposed by Heeger (1992) and has been studied extensively in the spatial domain, where the response r_i of neuron i is modelled as:

$$r_i(t) = K\frac{L_i(t)^n}{\sigma^n + \Sigma_k L_k^n(t)} \tag{4}$$

where K determines the maximum attainable response, n is an exponent, σ a semi-saturation constant and L_k^n the sum of the activations of neighbouring neurons. Here, we transition to the temporal domain and implement a mathematical framework originally formulated in Heeger (1992, 1993), which defines a neuron's response r_i recursively over time (**Fig. 2E**, *right*). For each unit i in the network the response at timepoint t is updated before applying the rectifier activation function ϕ such that:

$$r_i(t) = \phi\left(\lfloor L_i(t) \rfloor \frac{\sqrt{K - G_i(t-1)}}{\sigma}\right) \tag{5}$$

where $G(t-1)$ is a temporal feedback signal from the previous time step which is updated based on its previous state and the current response:

$$G_i(t) = (1-\alpha)G_i(t-1) + \alpha r(t-1) \tag{6}$$

where α determines the time scale of the feedback signal. This multiplicative feedback signal results in divisive suppression (for details, see Heeger 1993, Appendix A).

Recurrent interactions In addition to temporal adaptation arising from intrinsic biophysical mechanisms, temporal adaptation phenomena have also been proposed to be the result of recurrent interactions. Here, we incorporated temporal adaptation via lateral recurrence using a method adapted from Vinken et al. (2020), inspired by computational models which implemented adaptation by changing recurrent interactions between orientation tuned channels (Felsen et al., 2002; Teich and Qian, 2003; Westrick et al., 2016):

$$\mathbf{r}_n(t) = \phi\left(\mathbf{L}_n(t) + \mathbf{W}_n^l \mathbf{r}_n(t-1)\right) \tag{7}$$

where $\mathbf{W}_n^l$ are the lateral weights (**Fig. 2F**, *left*) consisting of 32 kernels of size 1 $\times$ 1 $\times$ 32 (stride = 1). To compare additive and multiplicative interactions, we also implement a form of lateral recurrence which feeds back activations using a multiplicative approach (**Fig. 2F**, *right*):

$$\mathbf{r}_n(t) = \phi\left(\mathbf{L}_n(t) \cdot \mathbf{W}_n^l \mathbf{r}_n(t-1)\right) \tag{8}$$

Implementation of contrast-dependent response modulation The relationship between stimulus contrast and the subsequent neural response has been empirically described by a contrast response function (CRF) following a sigmoid curve using the Naka-Rushton equation (Naka and Rushton, 1966):

$$r(c) = \frac{Ac^q}{c_{50}^q + c^q} \tag{9}$$

where r refers to the response given contrast c, A controls the saturation amplitude, q the steepness of the initial rise of the response and c_{50} the position of the response function along the contrast axis.

Here, we propose that in the DCNNs which apply temporal adaptation on the level of the single unit, namely by additive suppression and divisive normalization, we can incorporate this contrast-dependent modulation of neural responses, as described in Eq. 9, by varying the temporal adaptation parameters

which control the magnitude and shape of network activations, i.e. $[\alpha, \beta]$ for DCNNs with additive suppression and $[\alpha, \sigma, K]$ for DCNNs with divisive normalization (see **Fig. 2G** for the effect of varying α in DCNNs with additive suppression). Specifically, we implement contrast-dependent response modulation in two steps: We first train a network on full-contrast, noise-embedded digits during which the parameter values modulating the temporal adaptation are fitted simultaneously with the weights and biases (**Fig. 2H**, *left*). During inference, contrast-dependent response modulation is introduced by varying one of the temporal adaptation parameters, while keeping the other parameters controlling the temporal adaptation fixed (i.e. equal to the value obtained after training). We vary each temporal parameter based on the digit contrast level (**Fig. 2H**, *right*) using a modified version of Eq. 9:

$$\Theta(c) = \frac{Ac^q}{c_{50}^q + c^q} \tag{10}$$

where r is replaced by Θ which refers to the parameter value (e.g., α for additive suppression). By incorporating contrast in this way, we can test whether DCNNs as a result can capture the behavioral effect of both temporal adaptation and contrast gain. The motivation for varying this parameter value, instead of the network activations directly, originates from previous work, which has shown contrast affects the temporal dynamics of neural responses (Zhou et al., 2019; Groen et al., 2022). Replacing the response (r) in Eq. 9 with the value of the temporal adaptation parameter (Θ) in Eq. 10 accomplishes this, as it results in changes in the response shapes of the activations, including the transient and sustained response at stimulus on- and offset, respectively.

During inference, the parameters A, q and c_{50} were set as follows: A, representing the saturation amplitude, was set to the parameter value obtained after training (on 100% object contrast), q was set to 14 (however results were robust over different value settings) and c_{50} was varied across five different values, namely 0.5, 0.6, 0.7, 0.8 and 0.9 (corresponding to the object contrast levels, i.e. 50%, 60%, 70% , 80% and 90%). We then tested which setting of c_{50} best mimicked human performance, motivated by the fact that this parameter has been proposed to best capture contrast sensitivity across human observers (Baker and Graf, 2009). As mentioned above, the c_{50} determines the position of the CRF along the contrast axis and sets the contrast range over which responses are dynamic, and consequently to what degree the parameter value is adjusted for a given contrast level (**Fig. 2H**, *right*).

DCNN decoding analysis

To measure the amount of explicit information about the object present in DCNN activations and to compare with the object representations in the ERP responses, we also performed a decoding analysis on the DCNN unit activations of the last timestep (presentation of the test image) separately for each layer and contrast level. We applied a five-fold cross-validation, where each classifier was trained with 800 images (using the same type of classifier as was used for the decoding analysis of the ERPs, i.e. $\mathrm{sklearn.svm.SVC()}$) and tested on 200 images (all randomly drawn from the test set). This procedure was repeated five times after which accuracies were averaged over DCNN instances.

Data and code availability

All code used for the purpose of this paper can be found at the GitHub repository `https://github.com/ABra1993/tAdaptation_EEG.git`. The EEG data will be made publicly available on OpenNeuro when accepted for publication.

Results

To measure the joint effects of temporal adaptation and contrast gain on object recognition, we collected EEG data while participants categorized objects embedded in noise patterns, whereby we varied both a noise adapter preceding the object and the contrast level of the object (**Fig. 2A-C**). Below, we first present the perceptual effects of our manipulations, after which we show how adaptation and contrast affect event-related responses (ERPs). Then, we compare multiple computational mechanisms that could mediate these perceptual and neural effects using deep convolutional neural network (DCNN) modeling and investigate which networks best capture human behaviour. Lastly, we investigate the object-representations present in the evoked neural responses and DCNN activations by performing a decoding analysis and investigate potential benefits using either multiplicative over additive computations to enforce temporal adaptation.

Interactive effects of temporal adaptation and contrast gain on object recognition

Adding noise to the test images strongly affected object recognition accuracy: whereas subjects readily recognized clean objects presented without any noise (mean categorization performance = 98.15%, SD = 0.30%), performance dropped substantially when objects were embedded in noise patterns (mean categorization performance = 60.61%, SD = 9.39%), with notably stronger deterioration in performance as the contrast level decreased (**Fig. 3A**). We analyzed categorization accuracy across adapter types and contrast levels using a linear mixed effects model (see Materials and Methods, *Statistical testing for behavioral data*) that tested for effects of adapter type (blank, same and different) and contrast level (50%, 60%, 70%, 80% and 90%). We found a main effect of adapter ($F_{(2,4)}$ = 28.24, $p < 0.001$) and contrast ($F_{(2,4)}$ = 141.81, $p < 0.001$), as well as an interaction ($F_{(2,8)}$ = 2.75, p = 0.006) indicating that adaptation to noise differently affects performance across contrast levels. Adapting to noise had no effect for the two lowest contrast levels (50% and 60%), with similar performance for test images presented with and without adapter, while for contrast levels of 70%, performance improved after adapting both to same and different noise (same, $p < 0.001$; different, $p < 0.001$). For the highest contrast levels (80% and 90%), adapting to the same noise resulted in higher performance compared to no adapter (80% , $p < 0.001$; 90%, $p < 0.001$) or a different noise adapter (80% , p = 0.04; 90%, p = 0.0019). These results suggest that adaptation to same noise results in a stronger effect of contrast gain, compared to no or different-noise adaptation.

We also tested for potential interaction effects of adaptation and contrast gain on reaction times. Similar to the categorization accuracy, we found a main effect of adapter ($F_{(2,4)}$ = 105.09, $p < 0.001$) and contrast level ($F_{(2,4)}$ = 7.42, $p < 0.001$) as well as an interaction effect ($F_{(2,8)}$ = 2.81, p = 0.005).

When test images were presented with an adapter, response times were significantly higher compared to presentations without an adapter (for all object contrasts, pairwise $p < 0.001$) resulting in reaction times similar to test images without noise (**Fig. 3B**). Together, our behavioral analysis reveals a perceptual interaction effect between adapter type and contrast, with improvements in object recognition for high but not low object contrast levels as a result of adapting to preceding noise, accompanied by overall faster responses after adaptation.

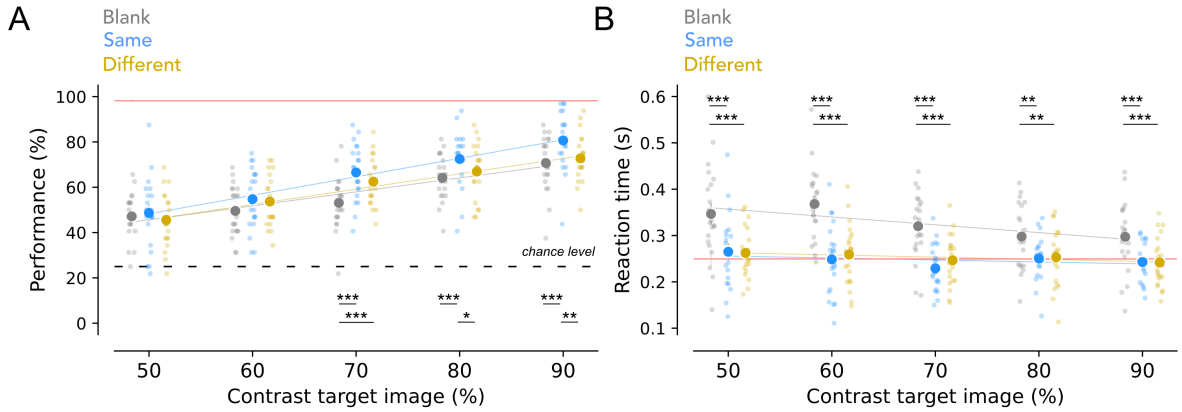

Figure 3: Interaction effect between adaptation and object contrast on categorization performance and reaction times. A: Recognition accuracy across contrast levels test images without adapter (grey) or with an adapter which contained the same (blue) or different (yellow) noise as the test image. The solid red line shows the performance for test images without noise; the dotted black line shows chance level (25%). Each point depicts an individual subject. **B**: Similar as (A) but for reaction times. Adaptation to same noise improves recognition performance for higher contrast levels, while adaption to both same and different noise results in faster reaction times. Linear mixed effect model, post-hoc Tukey test, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. This figure can be reproduced by mkFigure3.py.

Temporal adaptation affects early and late responses in occipito-parietal electrodes

We next examined how visually-evoked ERPs were modulated by either adaptation or object contrast level. All test images, regardless of presenting an adapter prior, elicited deflections in occipital and parietal electrodes early in the response (100-150 ms, **Fig. 4A**, *top*). Later in the response ($\sim$ 300 ms) ERP deflections were larger after adaptation (**Fig. 4A**, *bottom*). Differences were most pronounced in occipito-parietal electrodes, evident in a clear difference in ERP amplitude between the test images presented with and without adapters (**Fig. 4B**). Notably, same-noise trials show the largest deflection differences compared to blank trials. These results show that adaptation to noise causes diverging ERP deflections, with adaptation to same noise resulting in larger response magnitudes later in time.

To quantify these response differences across trial types and to gain more insights in the specific time windows during which adapter-related differences occur, we determined the time points at which

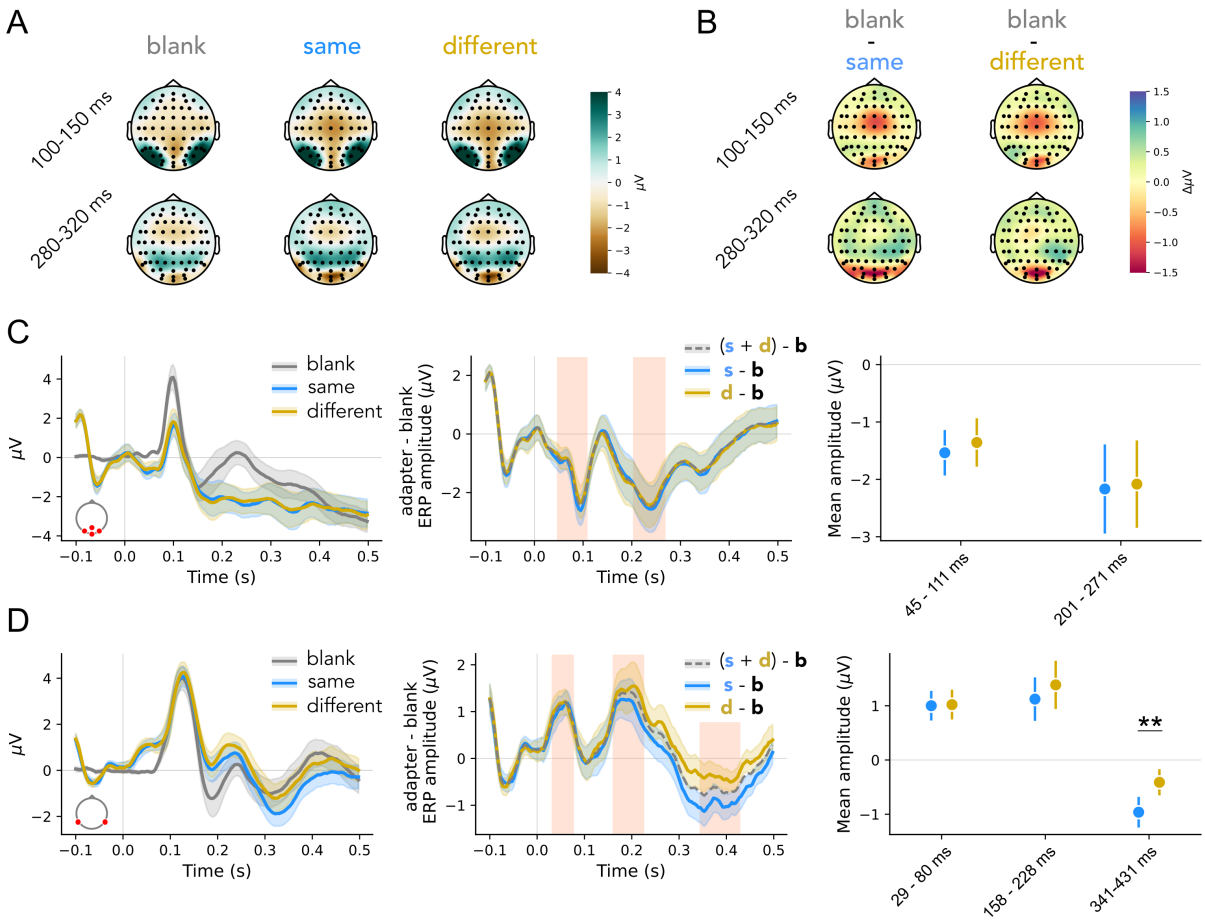

Figure 4: Temporal adaptation results in enhanced evoked activity at parieto-occipital electrodes. A: Topomaps representing ERP amplitudes for test images presented without adapters (blank, left) and test images presented after adapting to same (middle) or different noise (right). **B**: Difference in evoked potentials to test images with and without adapters (left, same noise versus blank; right, different noise versus blank). **C**: Left, average ERP per condition for electrodes Iz, Oz, O1 and O2. Middle, differences in ERPs with and without adaptation. Time windows for test images with adaptation for which same- (s) and different-noise (d) conditions are significantly different from blank (b) conditions are depicted by shaded red areas. Right, average ERP amplitudes for same vs. different noise adaptation within the identified time windows. Shaded regions and errorbars depict SEM across subjects. **D**: Same as C, but for electrodes P9 and P10. Adaptation-specific differences arise in parietal electrodes later in the response, with stronger deflections when adapting to same as opposed to different noise. T-test (two-sided), ** $p < 0.01$. This figure can be reproduced by mkFigure456.py.

responses with and without adapters differ significantly from one another (see Materials and Methods, *Identification of time windows of interest*). After identifying these time windows, we computed ERP response magnitudes separately for same- and different-noise adapters. For occipital electrodes we identify two significant adaptation-related time windows (45-111 ms and 201-271 ms), but within these windows, there was no significant difference between same and different adapters types (**Fig. 4C**). For parietal electrodes, we identify three time windows, whereby similar to occipital electrodes, the first two (29-80 ms and 158-228 ms) did not show differences across adapter types, but during the third time window (341-431 ms), same-noise adapters resulted in larger deflections compared to different noise adapters (dependent-samples T-test, $t_{(20)}$ = -3.54, p = 0.002, **Fig. 4D**). These results show that prior adaption to

noise affects neural responses to test images in both occipital and parietal electrodes, with adapter-specific effects emerging later in time at parietal electrodes, whereby same-noise adaptation results in stronger ERP deflections than different-noise adaptation, possibly reflecting facilitated neural processing of the object stimulus.

(Occipito)-parietal electrodes exhibit contrast-dependent response modulation

We have shown evoked responses are modulated by preceding inputs and that adaptation-specific differences occur later in the response. In addition to manipulating temporal adaptation to noise, we also varied the object contrast. We found no object contrast-dependent differences early in the ERP response ($<$ 200 ms, **Fig. 5A**), while later in time ($>$ 200 ms), responses in both occipito-parietal and parietal, but not occipital channels varied systematically with object contrast (**Fig. 5B-D**). To test for the effect of object contrast, we performed a linear fit on the late time window (see Materials and Methods, *Identification of time windows of interest*) and found a significant non-zero slope for parietal electrodes (one-sample T-test, slope of linear fit against 0, $t_{(20)}$ = -8.93, $p < 0.001$, **Fig. 5B**) and occipito-parietal electrodes ($t_{(20)}$ = 6.24, $p < 0.001$, **Fig. 5C**), but not for occipital electrodes ($t_{(20)}$ = -0.51, p = 0.62, **Fig. 5D**). These result show that response amplitudes of parietal-occipital electrodes, in addition to being modulated by previous inputs as we showed in the previous section, are also modulated by the contrast of the object.

Parietal electrodes show interaction of temporal adaptation and contrast at later time windows

So far, we found interaction effects between temporal adaptation and object contrast on categorization performance, as well as independent effects of these factors on ERPs. How do these two task manipulations jointly affect neural processing and how is this reflected in the evoked responses? Above we found that both adaptation to noise as well as varying contrast levels affect response magnitudes for parietal electrodes (P9 and P10). Therefore, we specifically focus on these electrodes to test for joint effects. Consistent with behavior, we find that parietal electrodes exhibit larger contrast-dependent modulation after adaptation, compared to the blank condition (**Fig. 6A**). Performing a linear regression on response magnitudes during the late time windows across object contrast levels shows that adapting to the same noise results in significant contrast-dependent modulation of the evoked responses (**Fig. 6B**), with higher magnitudes with increasing contrast levels (one-sample T-test, slope of linear fit against 0, $t_{(20)}$ = 3.32, $p = 0.003$). This effect however is not significant for different-noise trials ($t_{(20)}$ = 0.57, $p = 0.57$) or blank trials ($t_{(20)}$ = 0.62, $p = 0.54$). Moreover, these adapter-specific effects on contrast gain were not observed in the occipital and

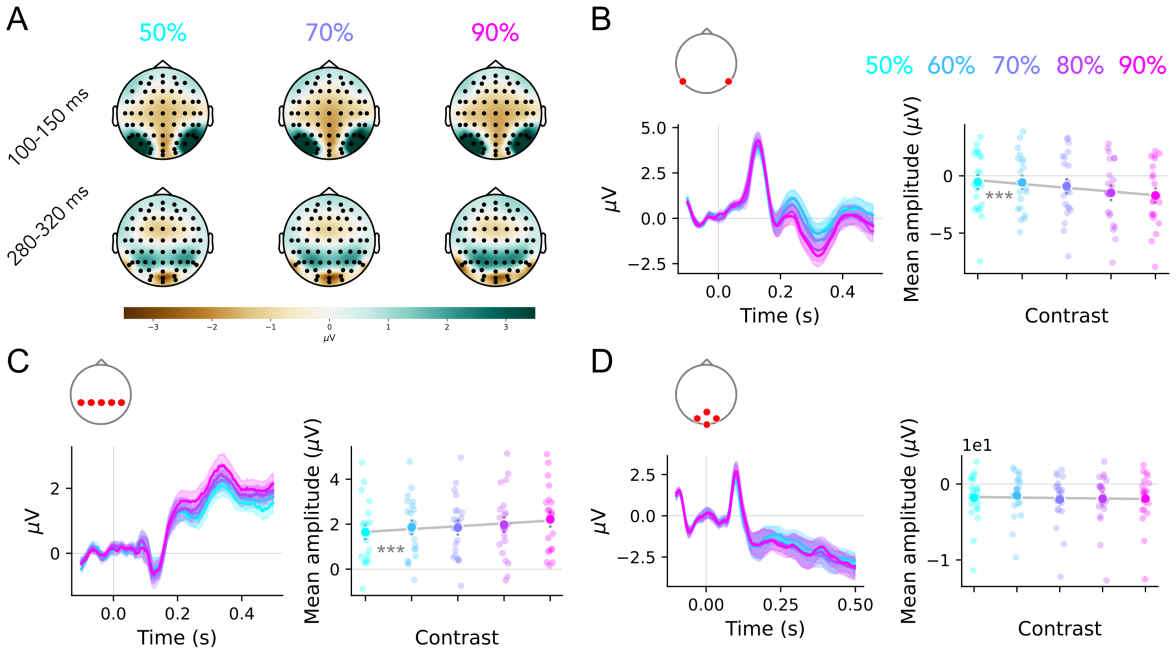

Figure 5: Responses in (occipito-)parietal electrodes modulated by contrast level. **A**: Topomaps representing ERPs during presentation of the test image for three object contrast levels. **B**: Left, ERPs shown separately per contrast level for P9 and P10. Right, response magnitude computed by taking the mean amplitude for the P300 component. **C-D**: Same as B for occipito-parietal (C) electrodes, including Pz, P1, P2, P3 and P4, and occipital electrodes (D), including Iz, Oz, O1 and O2. The P300 component of (occipito-)parietal electrodes is modulated by object contrast. 1-sample T-Test (coefficient of linear curve against 0), *** $p < 0.001$. This figure can be reproduced by mkFigure456.py.

occipito-parietal electrode groups. To conclude, our EEG data reveals a neural correlate of the perceptual effects at parieto-occipital electrodes, which show a more pronounced effect of contrast gain for same-noise as opposed to different-noise or blank trials.

DCNNs with temporal adaptation match human behavior for higher contrast levels

Our results so far show that temporal adaptation to same noise affects participants' ability to recognize objects, which is associated with stronger contrast gain of the neural responses. To discern potential computational mechanisms that could mediate these perceptual effects, we endowed DCNNs with four different adaptation mechanisms (**Fig. 2D-F**; see Materials and Methods, *Deep neural network implementation*). Two of these, divisive normalization and additive suppression, operate on the level of the single unit, whereby responsiveness to network inputs is reduced proportional to previous activations of only the unit itself. The other two adaptation mechanisms use lateral recurrence, operating across feature maps from previous timesteps and enforcing temporal adaptation through either additive or multiplicative interactions between units. After training on noise-embedded test images with full object contrast, DCNNs were tested on the same five contrast conditions as the human participants. We find that all four DCNNs generalize consider-

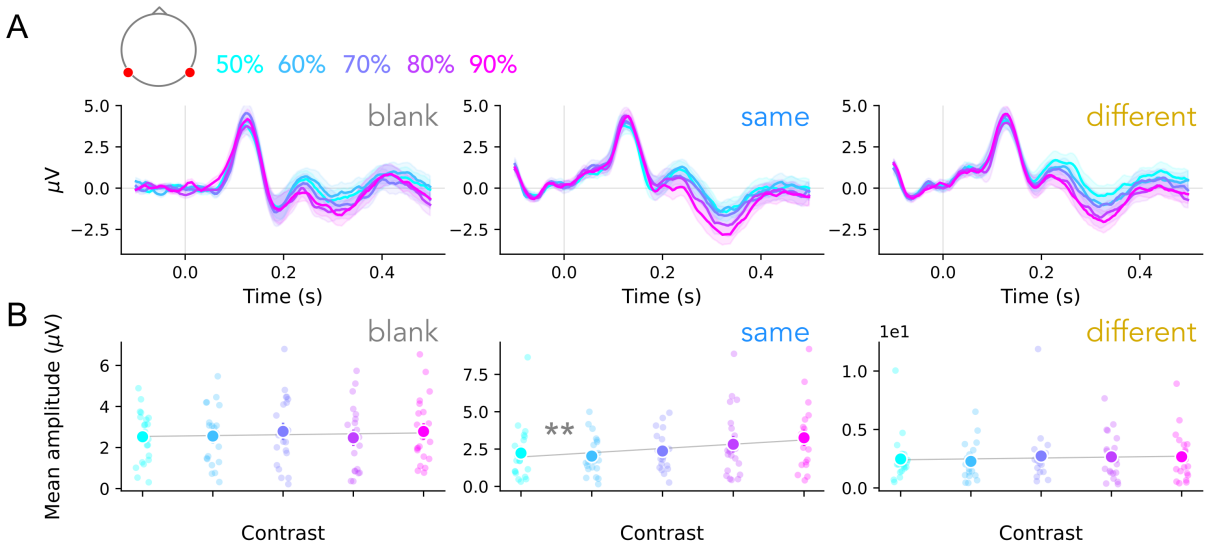

Figure 6: Joint effects of temporal adaptation and object contrast at parietal electrodes. A: ERPs for P9 and P10 shown separately for trials without adaptation (left) and adaptation with same (middle) or different (right) noise compared to the test image. The shaded regions depict the SEM across subjects. **B**: Response magnitude computed by taking the mean amplitude for the P300 component for the similar experimental conditions as in panel (A). When adapting to same noise, ERP signals exhibit contrast-dependent modulation evident by higher response magnitude with higher contrast levels. 1-sample T-Test (coefficient of linear curve against 0), ** $p < 0.01$. This figure can be reproduced by mkFigure456.py.

ably well to lower object contrasts, evident by the fact that performance is above chance level (25%) even for the lower contrast levels (**Fig. 7A**). Moreover, networks with temporal adaptation show a benefit of adapting to the same noise, evident by the fact that same-noise trials result in significant higher performance compared to different- and no noise trials (for all contrast levels, pairwise $p < 0.001$), in line with human behaviour. However, they fail to accurately predict the interaction effect between adapter type and contrast level, which in humans results in similar performance between same- and different-noise trials for the lower contrast levels. All DCNNs show the opposite pattern, namely a larger benefit of adaptation for lower rather than higher contrast levels. These results suggest that networks over-predict the benefit of temporal adaptation for lower object contrasts and do not correctly account for contrast gain, resulting in incorrect predictions for when the object contrast level is low.

DCNNs with variable responses capture interplay between adaptation and contrast

DCNNs with temporal adaptation mechanisms are able to capture human behavior for higher, but now lower contrast levels. This suggests that temporal adaptation mechanisms alone are insufficient to explain the interaction effect between adaptation and contrast gain observed in human subjects. To explain such interaction effects, networks require a mechanism that takes into account the contrast of the object. To achieve this, we propose to vary the parameters giving rise to temporal adaptation on the single unit level,

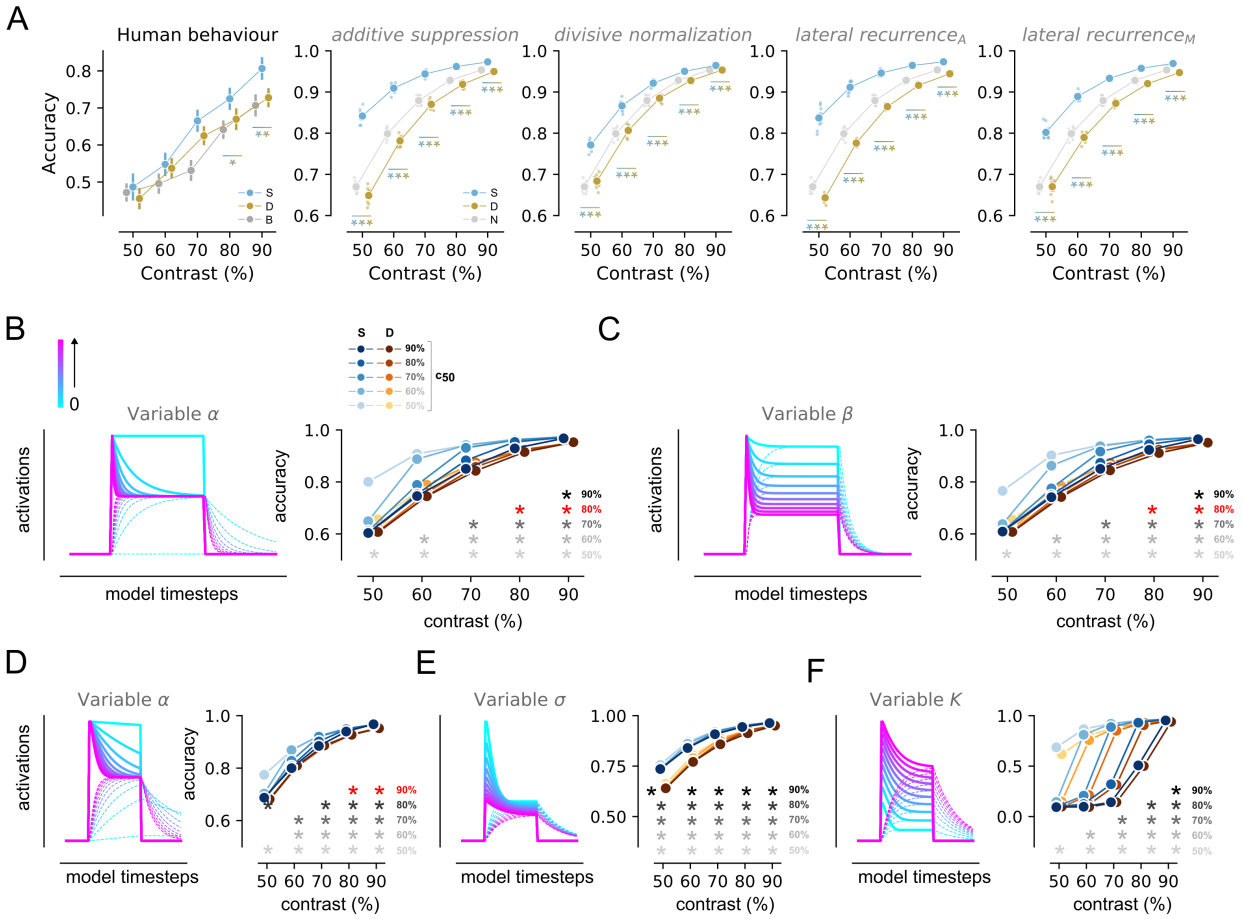

Figure 7: Interaction between temporal adaptation and contrast is captured by a DCNN with contrast-dependent response modulation. A: Left, Human object recognition performance for same- (blue, S), different- (yellow, D) and no noise (grey, B) trials (same data as in **Fig. 3A**). The remaining four panels show classification accuracy for the test set for same (blue, S) and different (yellow, D) noise adapters for DCNNs endowed with different temporal adaptation mechanisms (from left to right): additive suppression, divisive normalization, additive lateral recurrence (lateral recurrence$_A$) and multiplicative lateral recurrence (lateral recurrence$_M$). The grey markers show classification performance of networks without temporal adaptation (N) comparable with blank trials in humans. Error bars show SEM across network initializations ($n = 10$). T-test (two-sided) for same- and different-noise trials, * p $<$ 0.05, ** p $<$ 0.01, *** p $<$ 0.001. **B**: Effect of contrast-dependent modulation of network activations on categorization performance, by varying α for DCNNs with additive suppression. Left, Network activations for varying values of α. The solid lines depict the unit activations and the dotted lines depict the degree of temporal adaptation ($-\beta s(t)$ in Eq. 3). Right, Network performance for DCNNs with additive suppression with contrast-dependent modulation of α, separately for same (blue) or different (yellow) noise adapters for varying values of c_{50}. Grey asterisks denote whether network performance is significantly different (p $<$ 0.001) between same- and different-noise adapters. Network performance for the given c_{50} value which aligns with the human behavior (A) is shown in red (i.e. significant difference between same- and different-noise trials for a contrast of 80% and 90%). **C**: Same as (B) for the β parameter for DCNNs with additive suppression. **D-F**: Same as (B) for the parameters fitted for divisive normalization, namely α (D), σ (E) and K (F). For the left panels, the dotted line depicts the feedback signal for divisive normalization ($\frac{\sqrt{K-G_i(t-1)}}{\sigma}$ in Eq. 5). Adjustment of parameter values in a contrast-dependent fashion results in human-like network performance. Panel A can be reproduced by mkFigure7A.py and panels B-F can be reproduced by mkFigure7BCDEF.py.

namely α and β for additive suppression and α, σ and K for divisive normalization, in a contrast-dependent manner. These parameters control the magnitude and shape of the unit activations and thereby effectively determine the degree of adaptation over time (e.g., see **Fig. 2G** for varying the parameter α in a DCNN with additive suppression). The adjusted values of the temporal adaptation parameters are obtained based on a contrast response function (CRF; see Materials and Methods, *Implementation of contrast-dependent response modulation*), which is commonly used to characterize contrast responses in neural recordings (Naka and Rushton, 1966; Albrecht and Hamilton, 1982). We replace the response in the CRF with the value of the temporal adaptation parameter to explicitly model the effect of contrast on the adaptation parameters, which in turn affect unit response magnitudes. Note, the position of the CRF along the contrast axis (controlled by the parameter c_{50} in Eq. 10) determines the degree of contrast-dependent modulation (**Fig. 2H**, *right*): If the CRF is located left respective to the contrast axis, the contrast levels tested during inference are located in the saturating, right-side range of the CRF and therefore adjustment of the parameter value across object contrast levels will be weak. Consequently, the resulting network performance will be similar to the performances of the networks without contrast-dependent response modulation (**Fig. 7A**). On the other hand, when the CRF is shifted to the right respective to the contrast axis, thereby mimicking contrast gain, the dynamic contrast range will be higher and the parameter values (and consequently the unit activations) will be modified to a much stronger degree by the object contrast.

For DCNNs with additive suppression, varying the time scale of the decay (modulated by α, **Fig. 7B**) or varying the amount of suppression (modulated by β, **Fig. 7C**) using a CRF with a c_{50} of 80% captures the interaction between temporal adaptation and contrast gain, with little benefit of adaptation for lower contrast levels, as is also observed in human participants (**Fig. 7A**, *left*). For DCNNs with divisive normalization, varying the time scale of the decay (modulated by α, **Fig. 7D**) captures, similarly to DCNNs with additive suppression, the interplay, but requires a slightly stronger adjustment of the unit activations evident by the more rightward position of the CRF on the contrast axis (c_{50} = 90%). Varying the semi-saturation (modulated by σ, **Fig. 7E**) does not result in adjustment of unit activations such that the interplay between temporal adaptation and contrast gain is captured. Lastly, varying the maximal attainable response (controlled by K, **Fig. 7F**) does show the desired interaction effects. However, as the CRF moves along the contrast axis, network performance for lower contrast levels drops to chance level (i.e. 10%) which is not in line with the performance of human subjects, which exhibit above chance even for the lower contrast levels. While DCNNs with lateral recurrence could be modified such that object contrast could be accounted for, by for examples adjusting the weights, the convolutions used for this form of temporal adaptation are by nature linear and can therefore not capture, as opposed to DCNNs with additive

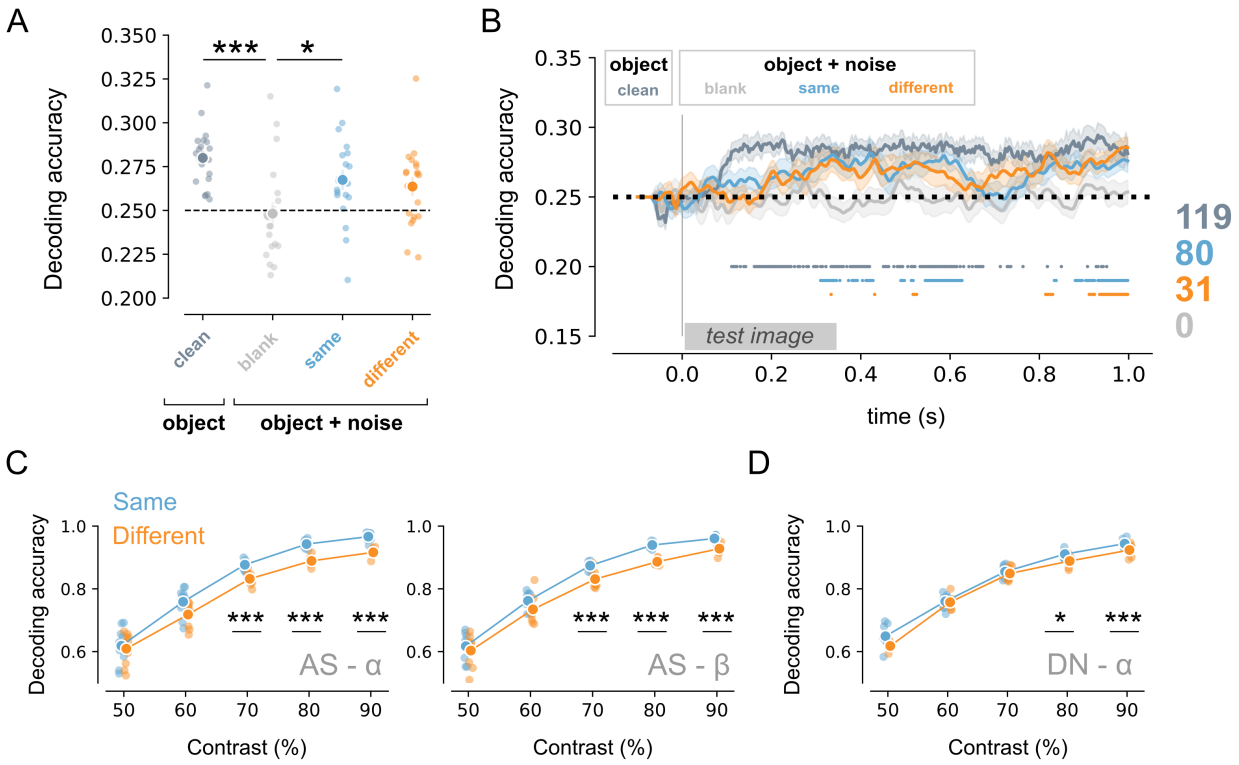

Figure 8: Object-related representations present in neural data and DCNN activations. **A**: Average decoding accuracy for predicting the presented object class based on evoked potentials for test images without (gray) and with embedded noise, including blank (light grey), same- (blue) and different-noise (orange) trials. Decoding accuracy's are averaged for [0, 1]s time window after stimulus onset. One-way ANOVA, post-hoc Tukey test, * $p < 0.05$, *** $p < 0.001$. **B**: Decoding accuracy for the different trial types across time points, with the number of time points for which decoding accuracy was significantly different from chance level (i.e. 25%) noted on the right. Shaded regions depict the SEM across subjects. **C**: Decoding accuracies for DCNNs with additive suppression and contrast-dependent modulation of either α (left) or β (right) computed over the activations of the fully connected layer for the CRF resulting in network performance similar to human participants (c_{50} = 80%). **D**: Same as (C) for DCNNs with divisive normalization and contrast-dependent modulation of α for the CRF resulting in network performance similar to human participants (c_{50} = 90%). Decoding accuracies for same-noise adaptation are higher compared to different-noise adaptation, which aligns with the neural data. Panel A-B can be reproduced by mkFigure8AB.ipynb and panel C-D can be reproduced by mkFigure8CD.py.

suppression and divisive normalization, the nonlinear changes due to object contrast. Altogether, these results reveal that DCNNs that take into account previous inputs and object contrast levels can account for the interaction effect of temporal adaptation and contrast gain by varying the parameters responsible for temporal adaptation in a contrast-dependent manner.

Object-representations of DCNNs align with neural data

Do DCNNs that align with human behavior and accurately capture the interaction effect between temporal adaptation and contrast gain also have similar object-representations as the neural data? To investigate this, we first determined to what extent object class can be decoded from the ERP responses (see Materials and Methods, *ERP Decoding analysis*). When objects are presented in full contrast without embedded

noise, decoding accuracy is significantly above chance (one-sample T-test, $t_{(20)}$ = 8.48, $p < 0.001$; **Fig. 8A**), suggesting that object-related representations are present in the evoked responses. Embedding the objects in noise heavily affects object-representations evident by the fact that for test images preceded by blank adapters, decoding accuracy drops to chance level. Remarkably, object-representations are partly restored when adapting to noise resulting in performance which is significantly above chance for both same- (one-sample T-test, $t_{(20)}$ = 3.36, p = 0.003) and different-noise adaptation (one-sample T-test, $t_{(20)}$ = 2.79, $p = 0.01$), implying the presence of object-related representations in these challenging viewing conditions is facilitated by temporal adaptation. Notably, we find a main effect of adapter type (one-way ANOVA, $F_{(3)}$ = 7.02, $p < 0.001$), with a significant difference in decoding performance between clean and blank trials ($p < 0.001$) and between same-noise and blank trials ($p = 0.036$), but not between different-noise and blank trials (p = 0.13), suggesting the presence of an adapter-specific effect on object-representations in the evoked activity. Indeed, decoding accuracy is significantly above chance for substantial more time-points during same-noise trials compared to different-noise trials (**Fig. 8B**), which aligns with the benefit of adaptation to same noise present on the perceptual level.

Next, we investigated whether DCNNs capture, analogous to the ERP decoding results, the more pronounced object-representations for same- compared to different-noise trials. DCNNs with additive suppression exhibit similar decoding patterns as the neural data: when we examine the DCNNs which exhibit human-like performance (c_{50} = 80%, **Fig. 7BC**), we find the networks show higher decoding accuracies for same- compared to different-noise trials, with decoding accuracies differing significantly for the higher contrast levels, namely 70% ($p < 0.001$), 80% ($p < 0.001$) and 90% ($p < 0.001$) object contrasts (**Fig. 8D**). DCNNs with divisive normalization also capture the difference in decoding accuracy between adapter types. When examining the networks for which the CRF results in human-like performance (c_{50} = 80%, **Fig. 7D**), we find significantly higher decoding accuracies with adaptation to same noise for 80% ($p = 0.003$) and 90% ($p < 0.039$) object contrasts (**Fig. 8D**). Overall, we show that object categories can be better recovered from ERP responses to noisy images as a result of temporal adaptation, and that DCNNs with temporal adaptation modulated by object contrast exhibit similar object representations as the neural data with higher decoding accuracies during same- as opposed to different-noise adaptation.

DCNNs with divisive normalization are more robust against spatially shifted noise

In the previous sections, we showed that DCNNs endowed with additive suppression and divisive normalization both capture human categorization performance and neural object-representations. Is there a

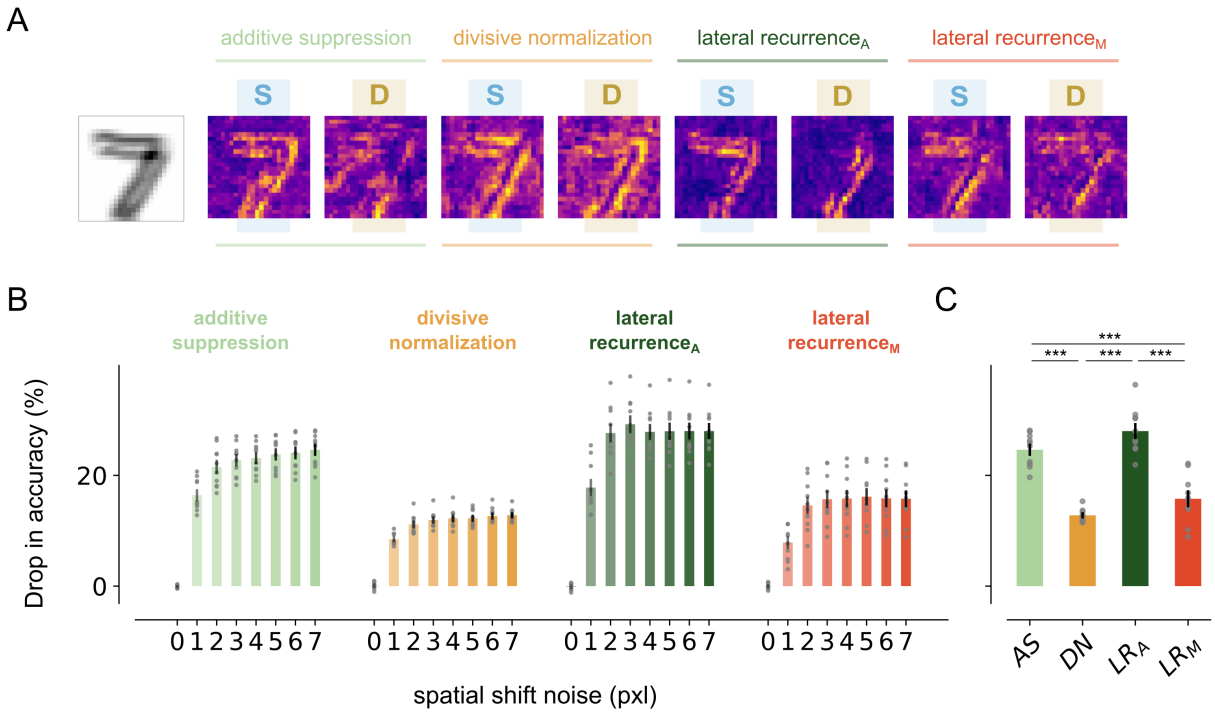

Figure 9: Networks using multiplicative computations are more robust against spatial shifts of input. A: Unit activations after the first convolutional layer at model timestep t_3 averaged over all feature maps for DCNNs with temporal adaptation. From left to right are outputs from DCNNs with additive suppression, divisive normalization, additive and multiplicative lateral recurrence for trials with same (S) or different (D) noise during adaptation and test. **B**: Effect of spatially shifting the noise during the test image on network performance. Depicted is the drop in accuracy compared to 0 pixel shift for the four temporal adaptation mechanisms. **C**: The average accuracy drops for the different spatial shifts depicted in panel B. DCNNs with divisive normalization (DN) and multiplicative lateral recurrence (LR_M) are more robust against spatially shifting noise during test time compared to DCNNs with additive suppression (AS) and additive lateral recurrence (LR_A). Individual points represent network initializations and error bars show the SEM. One-way ANOVA, post-hoc Tukey test, *** $p < 0.001$. Panel A can be reproduced by mkFigure9A.py and panel B by mkFigure9B.py.

possible advantage for biological systems to employ one form of temporal adaptation over the other? A key difference between additive suppression and divisive normalization is the interaction of the feedback modulation, namely additive suppression uses additive interactions to feed back activations, whereas divisive normalization uses multiplicative interactions. To further examine the effects of these interactions, we visualize the feature maps in the first convolutional layer. In addition to additive suppression and divisive normalization, we will also focus here on the two forms of lateral recurrence mentioned previously, which enforce temporal adaptation either by additive or multiplicative computations. DCNNs employing additive computations (i.e. additive suppression and additive lateral recurrence) exhibit stronger suppression of units that process pixels belonging to noise, compared to DCNNs employing multiplicative interactions (i.e. divisive normalization and multiplicative lateral recurrence; **Fig. 9A**). These results suggest that additive interactions result in selective suppression of noise units, whereas multiplicative interactions result in a broader suppression of all units regardless of the noise.

One possible explanation for the stronger noise suppression for DCNNs using additive interactions could be that these networks can directly subtract the noise, whereas DCNNs with multiplicative interactions cannot employ this strategy. We hypothesize that the fact that DCNNs cannot simply subtract the noise is beneficial, namely that unit activations are to some degree suppressed by the noise, but also still responsive to serve the classification task, which could possibly lead to better generalization. To test this hypothesis, we perform the following experiment: we use the same noise pattern for the adapter and test image, but for the latter we spatially shift the pixel values. The resulting test image has the same noise pattern but with a slight offset such that unit-wise subtraction is compromised. When shifting the noise pattern during test we see a drop in accuracy for all the temporal adaptation mechanisms (**Fig. 9B**). However, DCNNs employing multiplicative interactions, namely divisive normalization and lateral recurrence$_M$, are less affected by the spatially shifted inputs evident by the smaller drop in performance. In fact, we find a main effect (one-way ANOVA, F = 40.99, $p < 0.001$) with DCNNs employing multiplicative computations showing higher robustness compared to the DCNNs with additive temporal adaptation mechanisms (AS vs. DN, $p < 0.001$; AS vs. LR_M, $p < 0.001$; DN vs. LR_A, $p < 0.001$; LR_A vs. LR_M, $p < 0.001$; **Fig. 9C**). These results show that inductive biases involving multiplicative interactions increase robustness against spatially shifting the image property that induces temporal adaptation.

Discussion

Our aim was to examine the joint effects of two nonlinear neural response properties, temporal adaptation and contrast gain, on object recognition. We find an interaction effect with a benefit of adaptation to same noise for higher, but not lower contrast levels. This behavioral effect is associated with more pronounced contrast-dependent modulation of evoked responses after adaptation. We demonstrate that DCNNs implemented with temporal adaptation operating on the single-unit level allow natural incorporation of object contrast by adjusting temporal adaptation parameters using a biologically-inspired contrast response function. This modification allows DCNNs to capture the interaction effects on categorization performance and results in similar discriminability of object-representations as is observed in the neural data. Moreover, we reveal a benefit of temporal adaptation mechanisms using multiplicative computations, showing higher robustness against spatial shifts of the input.

The role of nonlinear response properties in perception

Nonlinear response properties have been implicated in robust perception. For instance, previous work has suggested that temporal adaptation could serve to decrease salience of recently seen stimuli (Solomon and Kohn, 2014; Vogels, 2016). We demonstrated this principle by testing human observers and DCNNs on a task with temporally repeated, but task-irrelevant noise patterns, whereby reducing the salience of recently seen features resulted in perceptual changes evident by increased categorization performance. In addition to temporal adaptation, we studied the effects of contrast gain, whereby pre-exposure to an adapter leads to perceptual constancy (Foley and Legge, 1981; Sclar and Freeman, 1982). Here we reveal a remarkable interaction between these two nonlinear response properties. That is, the benefit of temporal adaptation is only apparent with a sufficiently high contrast level, whereas the benefit of contrast gain is only observable when adapting to same noise. These findings emphasize the importance of studying such neural response properties simultaneously, as is the case in more naturalistic settings, as opposed to in isolation.

Signatures of nonlinear response properties in evoked neural responses

A large body of work has identified nonlinear response properties in EEG recordings, including temporal adaptation (Garrido et al., 2009; Engell and McCarthy, 2014) and contrast-dependent modulation (Schadow et al., 2007; Groen et al., 2013; Xi et al., 2020). In line with these findings, we revealed effects of temporal adaptation, with stronger deflections when adapting to same compared to different noise, and effects of contrast, with larger deflections as the object contrast level increases. More interestingly, we also observed

a joint effect of temporal adaptation and object contrast, with stronger contrast-dependent modulation when adapting to same compared to different or no noise. This finding suggests that the shift of the CRF due to pre-exposure to the adapter and the response reduction when adapting to same noise together result in a larger gain of the neural response, and more importantly, a stronger increase in categorization performance. An earlier study by Chaumon and Busch (2014) showed that a reduction of the gain in EEG responses results in reduced accuracy during a detection task, suggesting a link between the input-output-relationship and visual performance. Future research should investigate in more detail the possibly causal relationship between the gain in neural response magnitude and in categorization performance. Nonetheless, we show that the joint effects of temporal adaptation and contrast gain on visual performance are reflected in the evoked responses, revealing a link between neural processing and perception during object recognition.

Effective implementation of interactions between temporal adaptation and contrast gain in DCNNs

We show that DCNNs endowed with temporal adaptation operating on the level of the single unit capture the interaction between temporal adaptation and contrast gain on categorization performance, by modifying the temporal adaptation parameters in a contrast-dependent manner. This contrast-dependent modulation of individual units is supported by previous work which has shown CRFs adequately describe response profiles of single neurons (Albrecht and Hamilton, 1982; Geisler and Albrecht, 1992). The modified values for the temporal adaptation parameters were obtained using a CRF operating in a specific contrast range. This adjustment of network activations affects the initial response and the degree of suppression over time, which is in line with previous work showing the ratio between initial transient and later sustained responses in intracranial EEG responses are modulated by contrast (Zhou et al., 2019; Groen et al., 2022).

In addition to mechanisms operating on the single-unit level, we implemented two forms of recurrence operating across channels. While these networks capture the behavioral benefit of adaptation for higher contrast levels, they do not allow to naturally incorporate the object's contrast level by means of a pre-existing adaptation state parameter, and therefore fail to capture the interaction between temporal adaptation and contrast gain. These results may inform us about which perceptual effects could arise from visual processing within individual neurons and which arise due to recurrent interactions. Given the abundance of recurrence in the visual system (Felleman and Van Essen, 1991; Sporns and Zwi, 2004; Markov et al., 2014) it is most likely that temporal adaptation is implemented by a combination of intrinsic cell properties and recurrent connections. Future work could include other tasks or extend the current

model combining both intrinsic adaptation properties and lateral and/or top-down recurrence (Spoerer et al., 2017; Thorat et al., 2021; Lindsay et al., 2022). Nonetheless, we demonstrate how the role of nonlinear neural response properties in perceptual adaptation and object recognition can be elucidated using image-computable DCNNs, in line with the neuroconnectionist programme (Doerig et al., 2023).

Alignment of object-representations between neural data and DCNN activations

We show objects can be decoded from EEG responses, which is consistent with prior studies (Contini et al., 2017; Grootswagers et al., 2019). More interestingly, we observe an effect of temporal adaptation, where the drop in decoding accuracy when embedding objects in noise can be recovered by adapting to the same or a different noise pattern. While it has been previously shown that stimulus adaptation can improve decoding of objects from neural responses in monkey IT (Kaliukhovich et al., 2013), to our knowledge facilitatory effects of noise adaptation on object decoding from EEG signals have not been previously reported. We also show that adaptation to same noise results in higher decoding accuracies of the object class compared to adaptation to different noise, suggesting that the perceptual benefit of temporal adaptation may be mediated by improved discriminability of object-related neural representations. DCNNs with temporal adaptation modulated by object contrast show a similar pattern, with higher decoding accuracies from unit activations for same- compared to different-noise trials. Overall, we reveal a correspondence between informative features representing object identity in evoked responses and DCNN activations linking neural patterns to perceptual experience.

Multiplicative and additive interactions interactions

We offer a potential advantage for biological systems to use multiplicative computations: DCCNs with multiplicative interactions are more robust against spatially shifting inputs compared to networks that additively suppress activations. Previous work has similarly observed benefits of multiplicative interactions, showing improved object recognition of images consisting of composite images of multiple categories (Konkle and Alvarez, 2024). It has also been suggested that multiplicative effects play a role in shaping the CRF during attention-related processes (Liu et al., 2021) and are a useful inductive bias for conditional computations (Jayakumar et al., 2020). Overall, characterizing the unique role of multiplicative and additive interactions in how information flow is conditioned is a promising avenue and valuable addition to recent endeavors taken to explore network robustness (e.g. Geirhos et al. 2018).

Limitations and future work

First, during the object recognition task, object contrast was varied within a fixed range across subjects. Previous work however, has shown large heterogeneity in contrast sensitivity across individuals (Baker and Graf, 2009). Consequently, it is likely the lowest and highest contrast conditions yielded different neural response variations depending on subject-specific CRFs. Future studies could determine more fine-grained contrast ranges on the individual subject level, such that effects of temporal adaptation in dynamical as well as saturating contrast ranges can be investigated. Second, in the current study neural responses were collected with EEG. It is however unclear how CRFs of individual neurons translate to population responses. While in our EEG data analysis we quantify contrast-dependent modulation using a linear fit (**Fig. 6B**), previous studies have shown nonlinear IO relationships between visual inputs and neural responses (Chan et al., 2022) and is also what we assumed when determining the parameter value for the contrast-dependent response modulation in the DCNNs (**Fig. 2H**). To further investigate how contrast-dependent modulation affects neural responses and how this leads to perceptual experience in humans, other brain measurements could be deployed such as intracranial EEG, which can be measured more locally and whose high-frequency content is more closely related to spiking activity (Miller et al., 2009; Ray and Maunsell, 2011) and behavior (Miller et al., 2014). Lastly, there are several possible extensions to the current modeling framework, including testing on additional tasks (e.g. object occlusion, multiple-object recognition) and computational mechanisms (e.g. power law adaptation; Sörensen et al. 2023) to further study the role of various canonical computations and resulting response dynamics in robust object recognition.

Acknowledgements

This work was supported by a MacGillavry Fellowship to IIAG.

References

Albrecht, D. G. and Hamilton, D. B. (1982). Striate cortex of monkey and cat: contrast response function. *Journal of neurophysiology*, 48(1):217–237.

Baker, D. H. and Graf, E. W. (2009). On the relation between dichoptic masking and binocular rivalry. *Vision research*, 49(4):451–459.

Bates, D. M. and DebRoy, S. (2004). Linear mixed models and penalized least squares. *Journal of Multivariate Analysis*, 91(1):1–17.

Brands, A. M., Devore, S., Devinsky, O., Doyle, W., Flinker, A., Friedman, D., Dugan, P., Winawer, J., and Groen, I. I. A. (2024). Temporal dynamics of short-term neural adaptation across human visual cortex. *PLOS Computational Biology*, 20(5):e1012161.

Carandini, M. and Heeger, D. J. (2012). Normalization as a canonical neural computation. *Nature reviews neuroscience*, 13(1):51–62.

Chan, Y. M., Glarin, R., Moffat, B. A., Bode, S., and McKendrick, A. M. (2022). Relating the cortical visual contrast gain response to spectroscopy-measured excitatory and inhibitory metabolites in people who experience migraine. *Plos one*, 17(4):e0266130.

Chaumon, M. and Busch, N. A. (2014). Prestimulus neural oscillations inhibit visual perception via modulation of response gain. *Journal of Cognitive Neuroscience*, 26(11):2514–2529.

Contini, E. W., Wardle, S. G., and Carlson, T. A. (2017). Decoding the time-course of object recognition in the human brain: From visual features to categorical decisions. *Neuropsychologia*, 105:165–176.

Cravo, A. M., Rohenkohl, G., Wyart, V., and Nobre, A. C. (2013). Temporal expectation enhances contrast sensitivity by phase entrainment of low-frequency oscillations in visual cortex. *Journal of Neuroscience*, 33(9):4002–4010.

Dao, D. Y., Lu, Z.-L., and Dosher, B. A. (2006). Adaptation to sine-wave gratings selectively reduces the contrast gain of the adapted stimuli. *Journal of Vision*, 6(7):6–6.

del Mar Quiroga, M., Morris, A. P., and Krekelberg, B. (2016). Adaptation without plasticity. *Cell reports*, 17(1):58–68.

Di Russo, F., Spinelli, D., and Morrone, M. C. (2001). Automatic gain control contrast mechanisms are modulated by attention in humans: evidence from visual evoked potentials. *Vision research*, 41(19):2435–2447.

DiCarlo, J. J. and Maunsell, J. H. (2000). Form representation in monkey inferotemporal cortex is virtually unaltered by free viewing. *Nature neuroscience*, 3(8):814–821.

Doerig, A., Sommers, R. P., Seeliger, K., Richards, B., Ismael, J., Lindsay, G. W., Kording, K. P., Konkle, T., Van Gerven, M. A., Kriegeskorte, N., et al. (2023). The neuroconnectionist research programme. *Nature Reviews Neuroscience*, 24(7):431–450.

Engell, A. D. and McCarthy, G. (2014). Repetition suppression of face-selective evoked and induced eeg recorded from human cortex. *Human brain mapping*, 35(8):4155–4162.

Felleman, D. J. and Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral cortex (New York, NY: 1991)*, 1(1):1–47.

Felsen, G., Shen, Y.-s., Yao, H., Spor, G., Li, C., and Dan, Y. (2002). Dynamic modification of cortical orientation tuning mediated by recurrent connections. *Neuron*, 36(5):945–954.

Foley, J. M. and Legge, G. E. (1981). Contrast detection and near-threshold discrimination in human vision. *Vision research*, 21(7):1041–1053.

Garrido, M. I., Kilner, J. M., Kiebel, S. J., Stephan, K. E., Baldeweg, T., and Friston, K. J. (2009). Repetition suppression and plasticity in the human brain. *Neuroimage*, 48(1):269–279.

Geirhos, R., Rubisch, P., Michaelis, C., Bethge, M., Wichmann, F. A., and Brendel, W. (2018). Imagenet-trained cnns are biased towards texture; increasing shape bias improves accuracy and robustness. *arXiv preprint arXiv:1811.12231*.

Geisler, W. S. and Albrecht, D. G. (1992). Cortical neurons: isolation of contrast gain control. *Vision research*, 32(8):1409–1410.

Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., Goj, R., Jas, M., Brooks, T., Parkkonen, L., and Hämäläinen, M. S. (2013). MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroscience*, 7(267):1–13.

Gratton, G., Coles, M. G., and Donchin, E. (1983). A new method for off-line removal of ocular artifact. *Electroencephalography and clinical neurophysiology*, 55(4):468–484.

Grill-Spector, K., Henson, R., and Martin, A. (2006). Repetition and the brain: neural models of stimulus-specific effects. *Trends in cognitive sciences*, 10(1):14–23.

Groen, I. I., Ghebreab, S., Lamme, V. A., and Scholte, H. S. (2012). Spatially pooled contrast responses predict neural and perceptual similarity of naturalistic image categories.

Groen, I. I., Ghebreab, S., Prins, H., Lamme, V. A., and Scholte, H. S. (2013). From image statistics to scene gist: evoked neural activity reveals transition from low-level natural image structure to scene category. *Journal of Neuroscience*, 33(48):18814–18824.

Groen, I. I., Piantoni, G., Montenegro, S., Flinker, A., Devore, S., Devinsky, O., Doyle, W., Dugan, P., Friedman, D., Ramsey, N. F., et al. (2022). Temporal dynamics of neural responses in human visual cortex. *Journal of Neuroscience*, 42(40):7562–7580.

Grootswagers, T., Robinson, A. K., and Carlson, T. A. (2019). The representational dynamics of visual objects in rapid serial visual processing streams. *NeuroImage*, 188:668–679.

Gruber, T. and Müller, M. M. (2002). Effects of picture repetition on induced gamma band responses, evoked potentials, and phase synchrony in the human eeg. *Cognitive Brain Research*, 13(3):377–392.

Heeger, D. J. (1992). Normalization of cell responses in cat striate cortex. *Visual neuroscience*, 9(2):181–197.

Heeger, D. J. (1993). Modeling simple-cell direction selectivity with normalized, half-squared, linear operators. *Journal of neurophysiology*, 70(5):1885–1898.

Henson, R. and Rugg, M. (2003). Neural response suppression, haemodynamic repetition effects, and behavioural priming. *Neuropsychologia*, 41(3):263–270.

Jayakumar, S. M., Czarnecki, W. M., Menick, J., Schwarz, J., Rae, J., Osindero, S., Teh, Y. W., Harley, T., and Pascanu, R. (2020). Multiplicative interactions and where to find them.

Johnsdorf, M., Kisker, J., Gruber, T., and Schöne, B. (2023). Comparing encoding mechanisms in realistic virtual reality and conventional 2d laboratory settings: Event-related potentials in a repetition suppression paradigm. *Frontiers in Psychology*, 14:1051938.

Kaliukhovich, D. A., De Baene, W., and Vogels, R. (2013). Effect of adaptation on object representation accuracy in macaque inferior temporal cortex. *Journal of cognitive neuroscience*, 25(5):777–789.

Konkle, T. and Alvarez, G. (2024). Cognitive steering in deep neural networks via long-range modulatory feedback connections. *Advances in Neural Information Processing Systems*, 36.

LeCun, Y., Cortes, C., and Burges, C. (2010). Mnist handwritten digit database. *ATT Labs [Online]. Available: http://yann.lecun.com/exdb/mnist*, 2.

Lindsay, G. W., Mrsic-Flogel, T. D., and Sahani, M. (2022). Bio-inspired neural networks implement different recurrent visual processing strategies than task-trained ones do. *bioRxiv*, pages 2022–03.

Liu, C., Guo, F., Qian, C., Zhang, Z., Sun, K., Wang, D. J., He, S., and Zhang, P. (2021). Layer-dependent multiplicative effects of spatial attention on contrast responses in human early visual cortex. *Progress in Neurobiology*, 207:101897.

Lueschow, A., Miller, E. K., and Desimone, R. (1994). Inferior temporal mechanisms for invariant object recognition. *Cerebral cortex*, 4(5):523–531.

Markov, N. T., Vezoli, J., Chameau, P., Falchier, A., Quilodran, R., Huissoud, C., Lamy, C., Misery, P., Giroud, P., Ullman, S., et al. (2014). Anatomy of hierarchy: feedforward and feedback pathways in macaque visual cortex. *Journal of comparative neurology*, 522(1):225–259.

Mehrer, J., Spoerer, C. J., Kriegeskorte, N., and Kietzmann, T. C. (2020). Individual differences among deep neural network models. *Nature communications*, 11(1):5725.

Miller, K. J., Honey, C. J., Hermes, D., Rao, R. P., Ojemann, J. G., et al. (2014). Broadband changes in the cortical surface potential track activation of functionally diverse neuronal populations. *Neuroimage*, 85:711–720.

Miller, K. J., Sorensen, L. B., Ojemann, J. G., and Den Nijs, M. (2009). Power-law scaling in the brain surface electric potential. *PLoS computational biology*, 5(12):e1000609.

Naka, K.-I. and Rushton, W. A. (1966). S-potentials from luminosity units in the retina of fish (cyprinidae). *The Journal of physiology*, 185(3):587–599.

Nazari, M. A., Wallois, F., Aarabi, A., Nosratabadi, M., and Berquin, P. (2010). P300 component modulation during a go/nogo task in healthy children. *Basic and Clinical Neuroscience*, 2(1):31.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., et al. (2019). Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32.

Perrin, F., Bertrand, O., and Pernier, J. (1987). Scalp current density mapping: value and estimation from potential data. *IEEE Transactions on biomedical engineering*, (4):283–288.

Philiastides, M. G. and Sajda, P. (2006). Temporal characterization of the neural correlates of perceptual decision making in the human brain. *Cerebral cortex*, 16(4):509–518.

Ray, S. and Maunsell, J. H. (2011). Different origins of gamma rhythm and high-gamma activity in macaque visual cortex. *PLoS biology*, 9(4):e1000610.

Reith, F. H. and Wandell, B. A. (2020). A convolutional neural network reaches optimal sensitivity for detecting some, but not all, patterns. *IEEE Access*, 8:213522–213530.

Sawamura, H., Orban, G. A., and Vogels, R. (2006). Selectivity of neuronal adaptation does not match response selectivity: a single-cell study of the fmri adaptation paradigm. *Neuron*, 49(2):307–318.

Schadow, J., Lenz, D., Thaerig, S., Busch, N. A., Fründ, I., Rieger, J. W., and Herrmann, C. S. (2007). Stimulus intensity affects early sensory processing: visual contrast modulates evoked gamma-band activity in human eeg. *International Journal of Psychophysiology*, 66(1):28–36.

Schweinberger, S. R. and Neumann, M. F. (2016). Repetition effects in human erps to faces. *Cortex*, 80:141–153.

Sclar, G. and Freeman, R. (1982). Orientation selectivity in the cat's striate cortex is invariant with stimulus contrast. *Experimental brain research*, 46(3):457–461.

Solomon, S. G. and Kohn, A. (2014). Moving sensory adaptation beyond suppressive effects in single neurons. *Current biology*, 24(20):R1012–R1022.

Sörensen, L. K., Bohté, S. M., De Jong, D., Slagter, H. A., and Scholte, H. S. (2023). Mechanisms of human dynamic object recognition revealed by sequential deep neural networks. *PLOS Computational Biology*, 19(6):e1011169.

Spoerer, C. J., McClure, P., and Kriegeskorte, N. (2017). Recurrent convolutional neural networks: a better model of biological object recognition. *Frontiers in psychology*, 8:278016.

Sporns, O. and Zwi, J. D. (2004). The small world of the cerebral cortex. *Neuroinformatics*, 2:145–162.

Summerfield, C., Wyart, V., Johnen, V. M., and De Gardelle, V. (2011). Human scalp electroencephalography reveals that repetition suppression varies with expectation. *Frontiers in Human Neuroscience*, 5:67.

Teich, A. F. and Qian, N. (2003). Learning and adaptation in a recurrent model of v1 orientation selectivity. *Journal of Neurophysiology*, 89(4):2086–2100.

Thorat, S., Aldegheri, G., and Kietzmann, T. C. (2021). Category-orthogonal object features guide information processing in recurrent neural networks trained for object categorization. *arXiv preprint arXiv:2111.07898*.

Thorpe, S., Fize, D., and Marlot, C. (1996). Speed of processing in the human visual system. *nature*, 381(6582):520–522.

Vassilev, A., Stomonyakov, V., and Manahilov, V. (1994). Spatial-frequency specific contrast gain and flicker masking of human transient vep. *Vision research*, 34(7):863–872.

Vinken, K., Boix, X., and Kreiman, G. (2020). Incorporating intrinsic suppression in deep neural networks captures dynamics of adaptation in neurophysiology and perception. *Science Advances*, 6(42):eabd4205.

Vogels, R. (2016). Sources of adaptation of inferior temporal cortical responses. *Cortex*, 80:185–195.

Westrick, Z. M., Heeger, D. J., and Landy, M. S. (2016). Pattern adaptation and normalization reweighting. *Journal of Neuroscience*, 36(38):9805–9816.

Whitmire, C. J. and Stanley, G. B. (2016). Rapid sensory adaptation redux: a circuit perspective. *Neuron*, 92(2):298–315.

Wilson, H. R. and Humanski, R. (1993). Spatial frequency adaptation and contrast gain control. *Vision research*, 33(8):1133–1149.

Xi, J., Zhang, P., Jia, W.-L., Chen, N., Yang, J., Wang, G.-T., Dai, Y., Zhang, Y., and Huang, C.-B. (2020). Multi-stage cortical plasticity induced by visual contrast learning. *Frontiers in Neuroscience*, 14:555701.

Zhou, J., Benson, N. C., Kay, K., and Winawer, J. (2019). Predicting neuronal dynamics with a delayed gain control model. *PLoS computational biology*, 15(11):e1007484.