# Divisive normalization captures perceptual and neural interaction effects between temporal adaptation and contrast gain during object recognition

**Amber M. Brands (a.m.brands@uva.nl)**
Informatics Institute, University of Amsterdam
Science Park 900, 1098 XH, Amsterdam, The Netherlands

**Zilan Oz (zilan.oz@student.uva.nl)**
Informatics Institute, University of Amsterdam
Science Park 900, 1098 XH, Amsterdam, The Netherlands

**Iris I.A. Groen (i.i.a.groen@uva.nl)**
Informatics Institute, University of Amsterdam
Science Park 900, 1098 XH, Amsterdam, The Netherlands

Department of Psychology, University of Amsterdam
Nieuwe Achtergracht 129-B, 1018 WT, Amsterdam, The Netherlands

## Abstract

**Human perception remains robust under challenging viewing conditions. This robustness in perception has been linked to nonlinear processing of visual inputs. Here, we combine human EEG, behavior and deep neural network modeling to examine the joint impact of two nonlinear response properties, namely temporal adaptation and contrast gain, on perception of objects embedded in temporally repeated noise. We observe an interaction effect, with higher categorization performance when adapting to noise for high, but not lower object contrast levels. This improved performance is associated with more pronounced contrast-dependent modulation of the evoked neural responses and enhanced decoding of object identity. Using deep convolutional neural networks, we demonstrate that interaction effects between temporal adaptation and contrast level are effectively captured by temporal divisive normalization. Moreover, examining the network representations reveals that, similar to the neural data, adapting to the same noise results in improved representations of the object due to noise suppression. Overall, our findings shed light on how benefits of temporal adaptation are influenced by contrast level and offer an intuitive framework to study the integration of nonlinear response properties and their impact on perception.**



**Keywords:** temporal adaptation; contrast gain; object recognition; deep convolutional neural networks; divisive normalization


## Introduction

Our perception of sensory inputs depends heavily on nonlinear computations evident in various neural response properties, including temporal adaptation (reduced responses to repeating stimuli, **Fig. 1A**, *left*) and contrast gain (the sigmoidal relationship between stimulus contrast and the neural response, **Fig. 1A**, *right*). While these phenomena have each been studied extensively in isolation, their joint impact on perception, as well as the underlying computational mechanisms that give rise to neural responses and perceptual outcomes, is unclear. To study the joint impact of temporal adaptation and contrast gain on human visual perception, we collected neural and behavioural measurements while humans performed an object classification task with temporally repeated noise patterns, whereby the contrast of the object was varied. We endowed deep convolutional neural network (DCNN) with temporal adaptation mechanisms and evaluate their ability to predict human performance and EEG responses. We specifically compare temporal divisive normalization, a biophysically-realistic canonical computation that has been shown to capture nonlinear neural dynamics **(Heeger, 1992, 1993)**, with other types of temporal dynamics that have been introduced to DCNNs previously **(Vinken et al., 2020)**, namely an additive suppression mechanism and lateral recurrence.

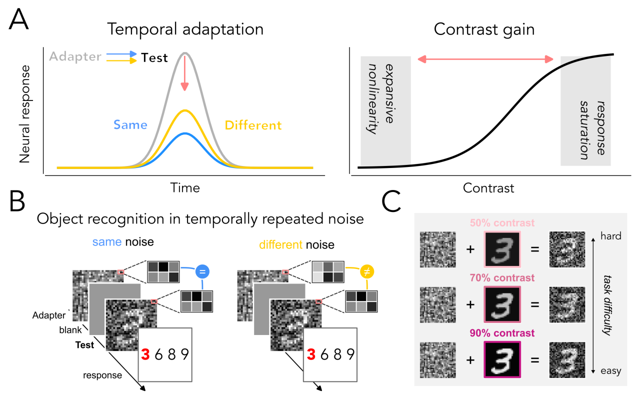


Figure 1: Experimental design. A: Nonlinear neural response properties. B: Classification task on objects (MNIST digits) in temporally repeated noise pattern. C: The contrast of the object is varied.

## Experimental procedure

### Data collection

Human subjects ($n = 21$) classified objects (MNIST, **Deng 2012**, classes *3*, *6*, *8* and *9*) embedded in pixelized noise patterns (Test, **Fig. 1B**) with varying contrast (**Fig. 1C**). Each Test stimulus was preceded by an Adapter, a noise-only pattern that consisted of either the same or different noise as presented during Test. Neural activity was measured using electrocorticography (EEG), from which we computed event-related potentials (ERPs) for a time window of [-100, 500] ms relative to the stimulus onset of the Test image and averaged trials within adapter types (i.e. same or different) and contrast levels, separately for each participant.

### Computational modeling

DCNNs were trained on the same task as performed by human participants. All DCNNs contained three convolutional layers, a fully connected and a readout layer, with one time step $t$ defined as one feedforward sweep. A training sample consisted of an image sequence, starting with the Adapter ($t_1$), a gray-scale image ($t_2$) and the Test ($t_3$). We define a linear response $\mathbf{L}_n(t)$ as the output of convolutional layer $n$:

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

given the unit's current input $\mathbf{x}_{n-1}(t)$, bottom-up convolutional weights $\mathbf{W}_n$ and biases $\mathbf{b}_n$. We compared four different temporal dynamics, two of which have been introduced previously **(Vinken et al., 2020)**, namely additive suppression ($AS$) and lateral recurrence. For the lateral recurrence we implement an additive ($LR_A$) and multiplicative ($LR_M$) form. In addition, we introduce a biologically-realistic implementation of temporal adaptation, known as divisive normalization ($DN$), inspired by previous work **(Heeger, 1992)**, as follows:

For each unit $i$ in the network the response is updated over time before applying the rectifier activation function $\phi$ so that

$$r_i(t) = \phi\left(\lfloor L_i(t) \rfloor \frac{\sqrt{K - G_i(t-1)}}{\sigma}\right) \quad (2)$$

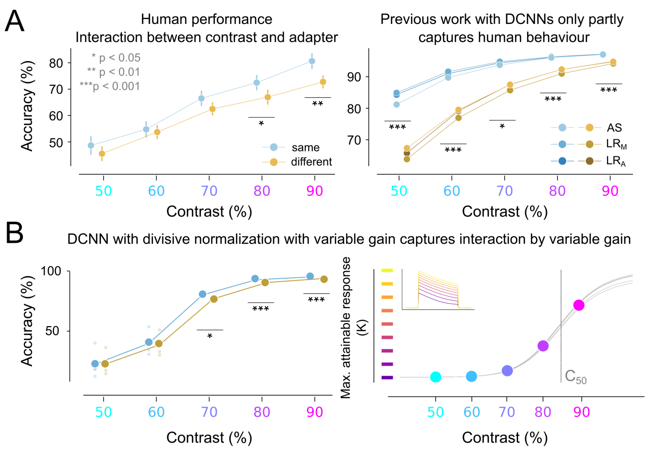


Figure 2: Human and DCNN recognition performance. A: Mean performance across human subjects for same (blue) and different (yellow) noise trials for each contrast level (left), and mean DCNN performance across multiple initializations ($n = 5$) (right). B: A DCNN with divisive normalization accurately captures human performance (left) by means of a variable gain mechanism (in the form of parameter $K$) which takes into account contrast (right).

where $K$ determines the maximum attainable response, $\sigma$ a semi-saturation constant and $G_i(t-1)$ a temporal feedback signal from the previous time step, which is updated based on its previous state and the previous response:

$$G_i(t) = (1-\alpha)G_i(t-1) + \alpha r_i(t-1) \quad (3)$$

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

## Results

**Interaction between contrast and adapter type is captured by a DCCN with divisive normalization.** In humans, we observe an interaction between temporal adaptation and contrast level: adapting to the same noise pattern improves accuracy for higher but not lower object contrasts (**Fig. 2A**, *left*). Training DCNNs with additive or lateral recurrent forms of temporal adaptation shows that network performances match human behaviour for higher but not lower contrast levels, with DCNNs showing the strongest benefits of adaptation for low rather than high object contrasts (**Fig. 2A**, *right*). However, a DCNN with divisive normalization better captures the interplay between the adapter type and contrast by allowing for the implementation of a variable gain mechanism (**Fig. 2B**, *left*). More specifically, we take into account contrast level by varying the maximal attainable value ($K$) for the first convolutional layer based on a contrast response function (CRF, **Fig. 2B**, *right*) following the Naka-Rushton equation (**Naka & Rushton 1966**). As a result, the adaptation benefits performance most in contrast ranges for which the responses are not plateauing. These results show that temporal adaptation is beneficial only for higher contrast levels, which is accurately captured by a DCNN with divisive normalization as an adaptation mechanism but not intrinsic suppression or lateral recurrence.

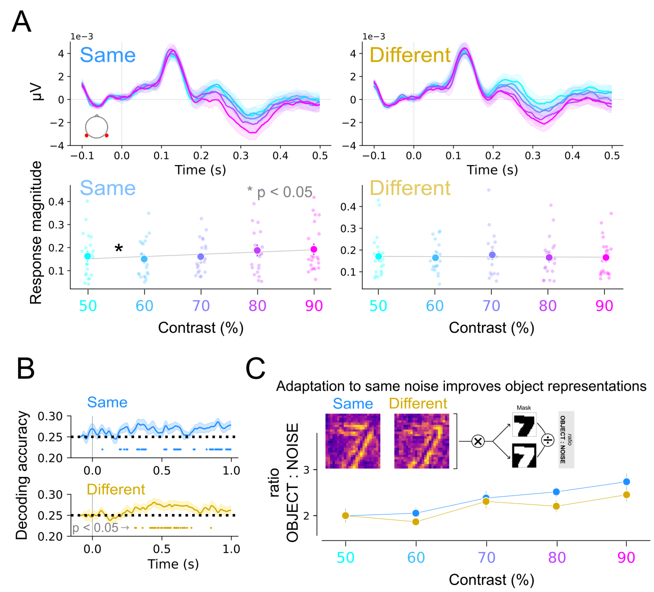


Figure 3: EEG responses and neural network activations. A: Top, ERPs to Test images at occipital-parietal electrodes for same (left) and different (right) noise trials per contrast level. Bottom, Response magnitude per contrast level computed by the Area Under the Curve. B: Object decoding accuracy performed on the ERPs of Test images after adaptation to same or different noise. C: Ratio of mean activations between DCNN units representing the object or the surrounding noise for averaged feature maps in the first convolutional layer.

**Benefit of temporal adaptation is mediated by improved representation of the object.** Analysis of the EEG data reveals that adaptation to the same noise results in more pronounced contrast-dependent modulation of the neural responses compared to adaptation to different noise (**Fig. 3A**). To examine whether adaptation influences neural object representations, we fitted a linear classifier predicting the digit class based on the ERPs evoked during presentation of the Test image. Results show that adapting to the same noise results in more prolonged significant decoding accuracy's (**Fig. 3B**), suggesting the object is more clearly represented in the neural signal as a result of adaptation. Examining DCNN activations shows that the behavioural benefit of adapting to the same noise is associated with improved representations of the object due to suppression of the surrounding noise (**Fig. 3C**). We quantify this by computing the ratio of mean activations between units representing the object and units representing the noise by applying masks, which shows that object representations for same noise trials shows a steeper increase compared to different noise trials as the contrast level increases. These results suggest that adapting to noise results in improved object representations in the brain, which matches improved representations reflected in DCNN activation patterns.

## Conclusion

Benefits of temporal adaptation on recognition of objects in noise depend on contrast level, which is accurately captured by a DCNN with temporal divisive normalization.

## Acknowledgments

This work was supported by a MacGillavry Fellowship to IIAG.

## References

Deng, L. (2012). The mnist database of handwritten digit images for machine learning research. *IEEE Signal Processing Magazine*, *29*(6), 141–142.

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.

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

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