



**HAL**  
open science

## Readout of the intrinsic and extrinsic properties of a stimulus from un-experienced neuronal activities: towards cognitive neuroprostheses.

Sabaa Farbod Kia, Elaine Astrand, Guilhem Ibos, S. Ben Hamed

### ► To cite this version:

Sabaa Farbod Kia, Elaine Astrand, Guilhem Ibos, S. Ben Hamed. Readout of the intrinsic and extrinsic properties of a stimulus from un-experienced neuronal activities: towards cognitive neuroprostheses.. *Journal of Physiology - Paris*, 2011, 105 (1-3), pp.115-122. 10.1016/j.jphysparis.2011.07.015 . hal-03082391

**HAL Id: hal-03082391**

**<https://hal.science/hal-03082391>**

Submitted on 28 Nov 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

# Readout of the intrinsic and extrinsic properties of a stimulus from un-experienced neuronal activities: towards cognitive neuroprostheses

Saba FARBOD KIA<sup>1,2</sup>, Elaine ÅSTRAND<sup>1</sup>, Guilhem IBOS<sup>1</sup>, Suliann BEN HAMED<sup>1,2</sup>

1. Centre de Neurosciences Cognitives, CNRS-UMR 5229, 67 Bd Pinel, 69675 Bron Cedex

2. Bahá'í Institute for Higher Education, [www.bihe.org](http://www.bihe.org)

Corresponding author: sbenha@isc.cnrs.fr

## ABSTRACT

While sensory and motor systems have attracted most of the research effort in Brain-Computer Interfaces (BCI), little attention has been devoted to higher order cortical processes [1]. Here, we propose to apply BCIs to the study and manipulation of visuospatial attention, an endogenous process at the interface between sensory and motor functions. As a first step to this aim, we investigate whether the activity of a population of frontal eye field neurons (FEF) in response to an endogenous cue can be readout *on a trial by trial basis* to provide a precise description of the cue's attributes, namely, its location and identity, but also the allocation of attention following its interpretation. Using the procedure described in Ben Hamed et al. [2, 3], we reach over 78% correct predictions for all decoded variables, including the spatial allocation of endogenous attention. We show that the decoding performance drops on incorrect trials, indicating that cue encoding participates to the animal's behavioral performance. Last, we show that the temporal resolution of the decoding influences readout performance. These results are a strong indication of the feasibility of the readout of endogenous variables by standard decoding algorithms, on a suboptimal dataset. Its validity remains to be proved in a real-time situation.

## KEY WORDS

Decoding, non-human primate, cortex, attention, perception, spikes.

## Introduction

While sensory and motor systems have attracted most of the research effort in BCIs [4, 5, 6; for review, 7, 8, 9], little attention has been devoted to higher order cortical processes [1]. From a clinical and societal point of view, such deficits of cognitive abilities are far from negligible and introducing BCIs to this field could lead to novel rehabilitation techniques. Here, we propose to apply BCIs to the study and manipulation of visuospatial attention, a process at the interface between sensory and motor functions. Endogenous spatial attention guides the active selection of sensory information based on cognitive priors; for example when looking for the phone, the likelihood of it being on the table is higher than on the ceiling. This process which is distinct from subsequent perceptual processes is a key cortical function. For example, following parietal acute lesions, hemineglect patients develop the

inability to attend and thus to perceive and interact with their contralesional environment [10, 11].

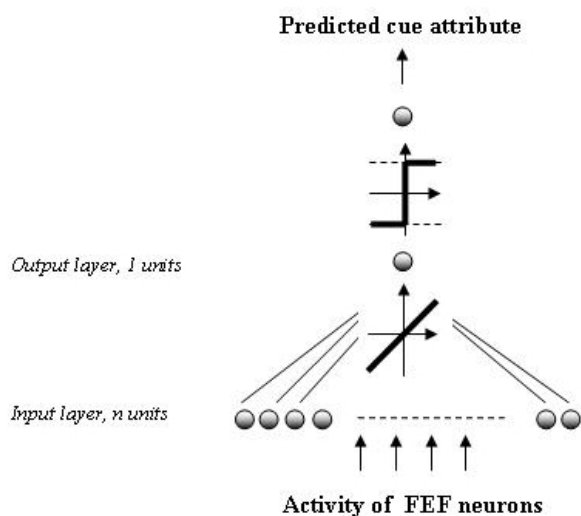
In a recent study, we have recorded from the FEF while monkeys were engaged in a variant of a cued target detection task [12, 13]. The experimental design of this task allowed to dissociate in time the processes related to the orientation of attention from processes related to target detection. In particular, a non-spatial abstract cue informed the monkey in which hemifield he should direct his attention. We found that FEF holds explicit information about the allocation of attention in response to the cue. We also showed that activity in this area reflects whether the monkey has perceived the target and is going to correctly detect it. We are currently investigating whether the information about attention allocation and perception can be read out from the activity of the FEF on a trial by trial basis, following the procedure described in [2] and [3]. We use a neural network approach which is formally equivalent to a linear decoding approach. Here we focus on the decoding of cue-related attributes (its location, its identity, and the allocation of attention following its interpretation) from the response of an FEF neuronal population. We provide an estimate of how the different attributes can be readout from *new neuronal single trial activities* and we discuss some of the factors that contribute to the improvement of this decoding.

## Methods

**Description of the neurophysiological database.** The activity of 123 frontal eye field (FEF) neurons were recorded from two macaque monkeys engaged in a cued target detection task (Ibos, Duhamel and Ben Hamed, 2009, submitted). The experimental design of this task allows to dissociate in time the processes related to the orientation of attention from those related to target detection. In particular, the cue is a non-spatial abstract cue that informs the monkey in which hemifield he should direct his attention. Briefly, the monkeys had to fixate a central point on the screen. Two streams of visual objects were presented, one in the visual receptive field of the neuron being recorded and the other in the contralateral side. One of the streams included a cue which instructed with a certain probability the position of the target. The cue could be green (resp. red), predicting that the target would appear in the same (resp. other) stream. The monkey had to release a lever to report the presence of the target. In 67% of the trials, the target appeared in the instructed stream (valid trials), in 17% of the trials, it appeared on

the opposite stream (invalid trials), and in 16% of the trials it didn't appear at all (catch trials) to discourage systematic responses. The monkey was rewarded for releasing the lever 150 to 750ms following target onset on valid and invalid trials and holding to it on catch trials. Invalid trials were used to check that the monkey used the predictive information provided by the cue in order to optimize his behavior. Sessions in which this was not the case were discarded from the analysis.

**Cell categories.** The recorded neurons were categorized as a function of their cue-related responses into different groups [12, 13]: neurons encoding the position of the cue (n=17), neurons encoding the instructed position of attention (n=20) and neurons specifically encoding one of the two types of cues (n=39). In all there were thus 76 cue related neurons.



**Figure 1.** Network architecture. The input layer encodes the activity of N FEF neurons following cue presentation (N=17, 20, 39 or 76, depending on the functional category of interest). The output layer contains 1 unit. The network is trained to encode the predicted attribute of the cue amongst two possible states, given current input activities. Training involves optimizing weights using a Levenberg-Marquardt back propagation algorithm and a linear transfer function. A symmetric hard limit function is applied to the 1 unit output in order to clip the readout to two discrete values, -1 and 1.

**Decoding.** We use a neural network approach which is formally equivalent to a linear decoding approach. The network was a one-layer feed-forward network constructed as follows (figure 1). The input layer of this network has one unit per cell of recorded in FEF during the whole length of the study, in the category of interest. Cells are thus artificially considered as one neuronal population. The output layer is a single binary unit corresponding to the state of one possible attribute of the cue: physical position (left/right), instructed position of attention (left/right) or cue identity (stay, shift). A linear activation function is used and the outputs of the network are clipped to two discrete values (-1 and 1) using a symmetrical hard limit function. *Training:* The input data is the cell activity averaged over 100 ms time bin centered around the time point where responses of interest have been shown to be maximal (200ms following cue or target onset). The network is trained on 50% of the trials available for each cell drawn randomly. Training is performed using Levenberg-Marquardt backpropagation function and the

weights are optimized to minimize the square distance between the estimate of the cue attribute and its true value. *Stopping:* Training is stopped using a regular stopping technique based on the performance of the network for a second subset of trials, called the validation set (random 20% of the remaining trials). *Testing:* The performance of the network is then tested for this time point (200ms post-cue) on the remaining trials, never experienced by the network, as well as on all trials for activities averaged over successive 100ms around the training time reference, with a 1ms resolution. The readout performance of the network is defined as the percentage of match between the output of the network and the actual physical state of the cue on the corresponding trials. This performance thus reflects whether the network can predict state of the cue when it is experiencing neuronal inputs that were collected when the monkey needed to interpret it for a successful behavior. The training testing procedure is repeated over 15 runs and yields an average decoding performance. This is the main measure that will be discussed in the following.

## Results

### Readout of the physical attributes of the cue.

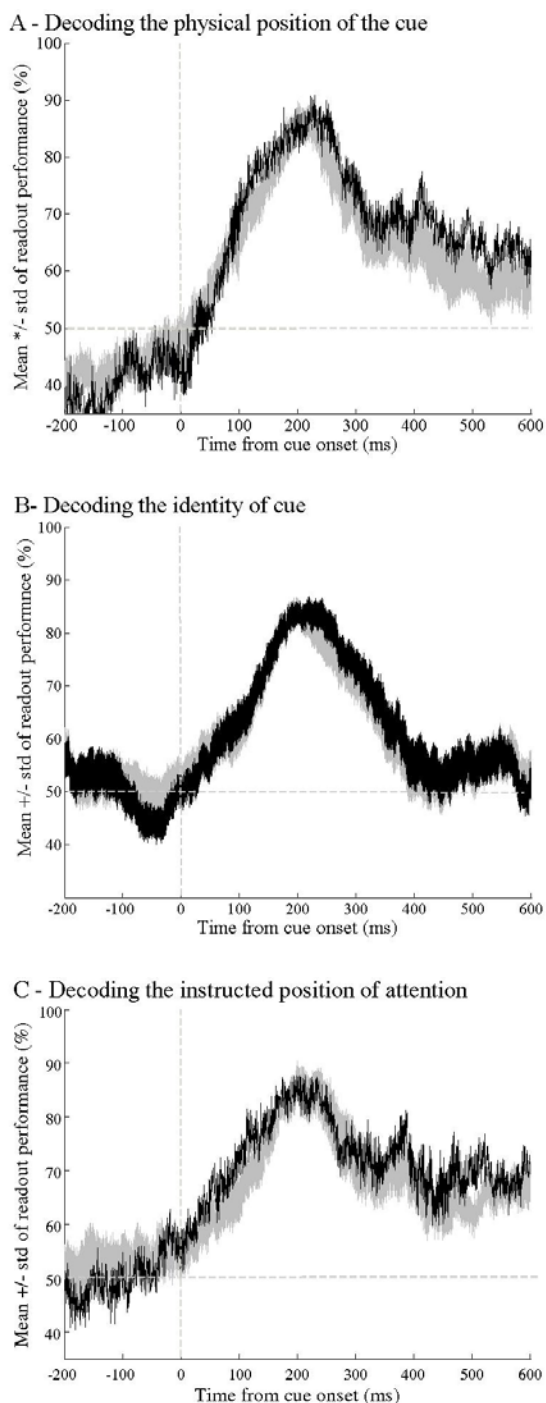
The first question that was addressed was whether the physical and abstract attributes of the cue could be readout from neuronal activities never previously experienced by the system. With a first network, we thus tried to predict the physical position of the cue, as described in the methods, either from the specific subpopulation of cells (fig.2a, black curve) or from the total cell population (fig.2a, gray curve). A maximum readout performance of 80% +/- 2 correct predictions is achieved in the 150-250ms time window following cue onset.

Later on, readout performance decreases, but remains above chance level (50%), at around 63% correct predictions. At this point in the task, other events are going on such as target presentation and detection, accounting for this decay in the availability of the information about the physical position of the cue.

With a second network, we sought to predict a non-spatial physical attribute of the cue, namely, its identity, whether the cue appeared in the left or in the right visual field (fig.2b, black curve, decoding from cells specifically encoding either the shift cue or the stay cue, gray curve, decoding from the total cell population). Here, a maximum readout performance of 86.1% +/- 2.3 correct predictions is achieved in the 150-250ms time window following cue onset. This readout performance then decreases, but stabilizes around 67%.

With a third and last network, we tried to readout the abstract instruction held by the cue, that is to say the final position of attention. Whereas cue position and cue identity refer to some physical property of the cue (position, or color), the instructed position of attention requires an interpretation of the cue. Indeed, while for the stay cue, the position of the cue is congruent to the instructed location of attention, for the shift cue, this is

not the case. The results are described in figure 2c (black curve, decoding from cells specifically encoding the position of attention, gray curve, decoding from the total cell population).



**Figure 2.** Readout performance of (A) the position of the cue, (B) the identity of the cue and (C) the instructed position of attention from the subpopulation selectively encoding the variable of interest (in black) and from the whole FEF population (in gray). Each curve represents mean +/- std of readout performance taken as the % of correct predictions of the network with respect to the actual configuration of the cue. The horizontal 50% performance line corresponds to random readout. The zero on the x-axis corresponds to cue onset. Note that on these data, the earliest timing at which a target could appear following the cue is 300ms.

Here, a maximum readout performance of 78.5% +/- 2.8 correct predictions is achieved in the 150-250ms

time window following cue onset. This readout performance then decreases, but stabilizes around 65%.

Interestingly, the readout performance was hardly affected by whether the decoding was performed on a selected subset of cells or on the whole population. This confirms the procedure by which neurons were classified in relation to their response to the cue. Indeed, this suggests that the connectivity weights assigned to the unit cells that are not contributing to the decoded variable are very low and do not participate to the variable being decoded. This is worth noting from a neuroprosthetic perspective, as the decoding is expected to be performed on all cells without a priori classification, the classification being carried out by the decoding procedure itself. As a consequence, in the following, all decodings will be performed on the whole FEF population.

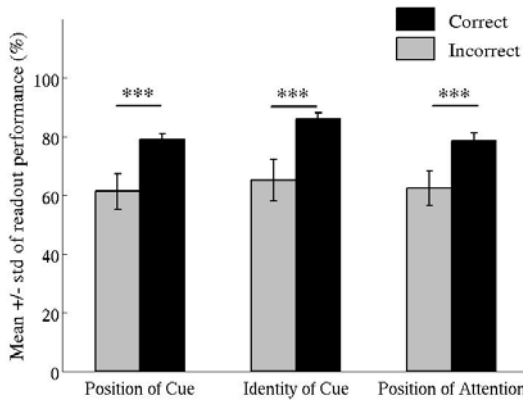
Not surprisingly the decoding performance is maximal at the delay used to train the network and in the surrounding 50ms bins (partially overlapping data). However, the achieved performance remains remarkable as it is calculated from neuronal response patterns that have not been experienced previously by the network. The decoding performance of the instructed position of attention is particularly noteworthy. Indeed, this attribute of the cue requires an endogenous operation leading to the interpretation of the cue from the combination of its physical attributes (position and identity).

#### **Effect of the animal's behavioural performance on the readout of the physical attributes of the cue.**

The task was designed so that the monkey needed to correctly interpret the cue in order to optimize its behavioral performance. Invalid trials allow us to check that this is indeed the case. Indeed, on these trials, the animals' reaction times are significantly slower and their performance significantly deteriorated with respect to valid trials, thus reflecting the behavioral cost of cue invalidity. This proves that the cue is used by the monkeys in order to perform the task. The question we are asking here is whether the encoding of the cue as reflected by the population response is different on incorrect trials than on correct trials.

We thus presented the networks trained to decode the different cue parameters (position, identity and instruction) with the population response on incorrect trials and compared the performance to that obtained for correct trials (figure 3). The readout performance for the position of the cue (80% +/- 2 vs. 61.4% +/- 6.1,  $p < 0.0001$ ), the identity of the cue (86.1% +/- 2.3 vs. 65.2% +/- 7.1,  $p < 0.0001$ ) and the position of attention (78.5% +/- 2.8 vs. 62.4% +/- 6,  $p < 0.0001$ ) are significantly deteriorated on incorrect trials with respect to correct trials, though they remain significantly different from chance ( $p < 0.0001$ ). This suggests that a miss-encoding of the cue has a partial causal relationship with the animal's behavioral performance. Obviously, the encoding of the target is also expected to have a causal effect on behavioral performance, error

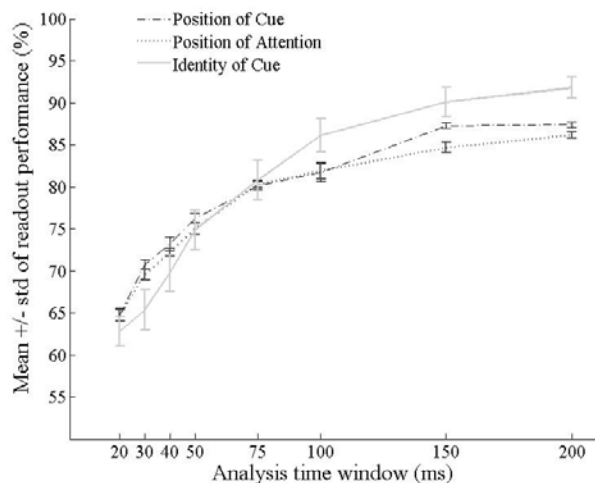
trials arising from a conjunction of miss-encoding of both the cue and the target.



**Figure 3.** Readout performance as a function of the animal's behavioural performance. In Black, decoding performance of the position of the cue, its identity and the instructed position of attention, on trials on which the monkey successfully performed the task. In gray, decoding performance on trials on which the monkey failed to correctly report the target's presence. \*\*\*:  $p < 0.0001$ .

### Effect of temporal resolution on the readout of the physical attributes of the cue.

Although the decoding is performed on 1ms time-steps, the activity is averaged over 100ms time bins. This has as effect to reduce noise and to smooth the neuronal responses. Here, we compare the readout performance of the different attributes of the cue, obtained for activities averaged over different window lengths (figure 4). We show that decoding improves steadily as bin width increases although this improvement stabilizes for analysis time windows beyond 100ms.



**Figure 4.** Readout performance as a function of neuronal activities time averaging window length.

While this difference in decoding performance between small and larger bins can seem surprisingly important, it is worth noting that the information that is being

decoded here is not a sensory information but rather a cognitive information involving an endogenous processing of the cue. Indeed, the cue is embedded into a visual stream of stimuli. We have shown elsewhere [13] that the FEF does not encode the transient changes in the stream of visual stimuli and that all the observed response modulations (apart from the response to stream onset), both at the single cell and at the population level were due to the cognitive processing of either the cue or the target. While sensory responses are expected to be locked to the stimulus onset, cognitive processes are expected to have a more variable temporal dynamics. Corroborating this point, the cue responses have an overall higher onset variability (std = 18.75) than the visual response to the first stream onset (std = 15.76),  $p < 0.000001$  on an F-test for variance equality. This might explain in part the results presented in fig. 4.

### Processing time

Neuroprosthetics requires a high read out performance of neuronal content achieved fast enough to allow a reactivity that lies in the biological range. Indeed, among the objectives of neuroprosthetics is the ability to predict an upcoming behaviour before it takes place, to produce a surrogate response to the naturally occurring one or to interfere with an ongoing cognitive process. This implies a trade-off between the improved decoding performance on larger time averaging windows and the incurred decrease in temporal resolution.

After an initial training of the decoding network on an average performance PC (Intel Xeon Quad Core E5430 bi-processor, 2.66 GHz, a TYAN S5397 mother board, 16 Go of RAM –DDR2-667 FB Dimm-, a RAID 3WARE 9550SX controller card and 6 hard discs SATA2 80 GO – 7200 cycle/min), we evaluated the average time taken by the readout on a given trial both for the total population ( $n=76$ ) and for the attention only cells ( $n=20$ ). This decoding time depended on the size of the population and was longest for the largest population (900ms versus 490,  $n=10$ ,  $p < 0.0001$ ).

### Discussion and conclusion

As a first step to reading out endogenous variables from cortical neuronal ensembles, we have hereby demonstrated the feasibility of decoding different attributes of a visual cue and most notably of its spatial attention significance with a performance above 78%, offline from neuronal activities collected separately. This performance is probably a lower limit on what can be achieved with real-time data; indeed, the neuronal activities presented to the network on given 'trial' were actually recorded independently. This is obviously a source of noise expected to draw this decoding performance to the worst, as each trial is associated with a different specific attentional dynamics. However, these preliminary data were free of the detrimental (in terms of information capacity) neuronal correlations that we expect to find in parallel recordings. Decoding endogenous variables from a neuronal population will be the next step in the present endeavor.

From a neuroscience point of view, our intuition is that the access to whole population dynamics in cortical oculomotor areas such as the FEF is a key tool to bridge the gap between the neuropsychology of visuospatial attention and the electrophysiological descriptions of individual cells. From a neurocomputational point of view, we believe that accessing to such decision-making processes will help refine the decoding of motor intentions, improve neuroprostheses and ultimately lead to novel rehabilitation techniques.

## Acknowledgements

The authors thank Drs Pierre Baraduc, Claire Wardak and the anonymous reviewers for constructive feedback on the present work and Sylvain Maurin for technical assistance. This project is funded by ANR-05-JCJC-0230-01 and CNRS 2009 Neuroinformatic grant.

## References

- [1] RA Andersen, JW Burdick, S Musallam, B Pesaran, JG Cham. Cognitive neural prosthetics. *Trends Cogn. Sci.*, 8(11), 2004, 486-493.
- [2] S Ben Hamed, W Page, C Duffy, A Pouget. MSTd neuronal basis functions for the population encoding of heading direction. *J Neurophysiol.*, 90(2), 2003, 549-58.
- [3] SB Hamed, MH Schieber, A Pouget. Decoding M1 neurons during multiple finger movements. *J Neurophysiol.*, 98(1), 2007, 327-33.
- [4] J Wessberg, CR Stambaugh, JD Kralik, et al. Real-time prediction of hand trajectory by ensembles of cortical neurons in primates. *Nature*, 408(6810), 2000, 361-365.
- [5] MD Serruya, NG Hatsopoulos, L Paninski, MR Fellows, JP Donoghue. Instant neural control of a movement signal. *Nature*, 416(6877), 2002, 141-142.
- [6] LR Hochberg, MD Serruya, GM Friehs, et al. Neuronal ensemble control of prosthetic devices by a human with tetraplegia. *Nature*, 442(7099), 2006, 164-171.
- [7] NG Hatsopoulos, JP Donoghue. The science of neural interface systems. *Annu. Rev. Neurosci.*, 2009, 32:249-266.
- [8] MA Lebedev, MAL Nicolelis. Brain-machine interfaces: past, present and future. *Trends Neurosci.*, 29(9), 2006, 536-546.
- [9] MAL Nicolelis, MA Lebedev. Principles of neural ensemble physiology underlying the operation of brain-machine interfaces. *Nat. Rev. Neurosci.*, 10(7), 2009, 530-540.
- [10] M Husain, P Nachev. Space and the parietal cortex. *Trends Cogn. Sci.*, 11(1), 2007, 30-36.
- [11] MJ Riddoch, M Chechlacz, C Mevorach, et al. The neural mechanisms of visual selection: the view from neuropsychology. *Ann. N. Y. Acad. Sci.*, 1191(1), 2010, 156-181.
- [12] G Ibos, JR Duhamel, S Ben Hamed. Differential contributions of parietal and frontal cortices to perceptual decision and attention control. *Society for Neuroscience abstract*, Chicago, 2009.
- [13] G Ibos, JR Duhamel, S Ben Hamed. Differential contributions of parietal and frontal cortices to perception and attention control. *Submitted*.