Neural Control: Closed-Loop Human Brain Reading

Closed-loop experimental testing of single medial temporal lobe neurons in humans reveals top-down effects, opening new possibilities for describing neural representations at the highest level.

Peter Földiák

Functional imaging methods give us some information about computations in the brain based on the averaged activity of millions of neurons. Knowing when and where such activity happens, however, is far from knowing what is computed and how. It is like listening to a busy conference poster session from a distance: you may know that some regions are busier than others, but you do not learn much from the cacophony unless you move closer and listen to individual presenters. Electrodes surgically implanted in the brains of human patients for the treatment of epilepsy make it possible to record single neurons and small groups of neurons, which lead to the discovery of a highly explicit, narrowly tuned neural representation of concepts in the medial temporal lobe (MTL, including the amygdala, enthorinal cortex, parahippocampal cortex and hippocampus) [1].

In a recent example of such experiments, Cerf *et al.* [2] focused on the voluntary, 'top-down' control of these MTL neurons using an innovative closed-loop, on-line experimental design, rarely used even in animal experiments (Figure 1). They first selected four neural units for recording and determined their preferred stimuli (for example, images of familiar faces). The preferred images of two of the four recorded neurons were then used to make a superimposed hybrid stimulus. The patient was able to change the mixing proportions of the two images during the experimental trial by their own neural activity. The patient's task was to make a target image dominate the stimulus, fading out the other image. Interestingly, even when patients were not told how to do this, they often succeeded right from the beginning of the experiment, probably by paying attention to the target. At each time step during the trial, the mixing ratio was changed in the direction of the image for which the previously measured neural activity pattern was closer to the activity pattern measured at that moment.

The results of Cerf *et al.*[2] raise a number of questions. For example: do we control our brain? The question of the voluntary control of our brain activity in general is only superficially interesting, as it assumes a false, dualistic separation of 'we' and 'our brain activity'. The more interesting question is over which neurons in which areas can top-down, cognitive intervention exert control. Such control is obvious in primary motor cortex, which drives our voluntary movements. Also, there is no reason why such voluntary influence should be limited to primary motor cortex, and it is conceivable that we have some kinds of 'virtual arms' or switching circuits that can be operated inside our brains at will. After all, humans can take instructions from experimenters and set up their brains within seconds to perform a wide variety of tasks voluntarily. In the case of the MTL, the question of 'top-down' influence is interesting as these areas are at the top of the sensory processing hierarchy. One possibility is that these top-down effects come from areas in the executive system, such as prefrontal cortex, but here the question of which area is 'top' and which is 'down' is no longer so clear.

Is this evidence for strong voluntary control? It has long been known that neural activity is not only determined by external stimuli but also modulated by 'top-down' influences such as context, attention and the other forms of internal state of the brain [3,4]. Neural connections from higher to lower areas are also prominent anatomically. Therefore, it is interesting, though not surprising, that such top-down effects can also be observed in human MTL. Not all top-down effects are necessarily voluntary though, and Cerf *et al.* [2] do not prove that the observed effects are truly voluntary (for example, repetition priming has not been ruled out). Also, the magnitude of the top-down effect is impossible to determine from these experiments, as the positive feedback may magnify small initial biases through changes in the stimuli. The competition seen in these experiments could be due to the experimental loop, and it is not a proof of competition in sensory processing itself (Figure 2). This positive feedback gives us an easily visible outcome for each trial, but an alternative, negative feedback loop design may give us a measure of the strength of the top-down effect.

Is the closed-loop design useful for describing neural selectivity? The traditional method of presenting a fixed set of stimuli selected based on the expectations, intuitions and biases of the experimenter (usually varying in some simple set of parameters) and measuring the neural tuning across this set is highly restrictive. For high-level areas, however, the parameters themselves are highly complex and hard or impossible to guess. Various forms of unstructured or structured noise or natural stimuli may be used for correlation [5,6] or system identification [7], but these methods are unsuitable for the complex nonlinearities introduced by many layers of sensory processing. Closed-loop methods may be useful in these cases, as they allow the selection of the testing stimuli to depend on previous responses. Imagine the difference between a normal game of 'twenty questions', where a word is guessed based on the yes or no answers to a series of questions, and one where the questions must be fixed in advance. Surely the closed-loop version of this game is orders of magnitude quicker than the fixed questions version, as the previous answers eliminate large parts of the space of possibilities.

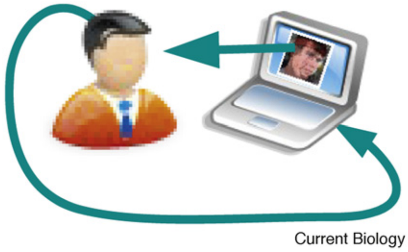
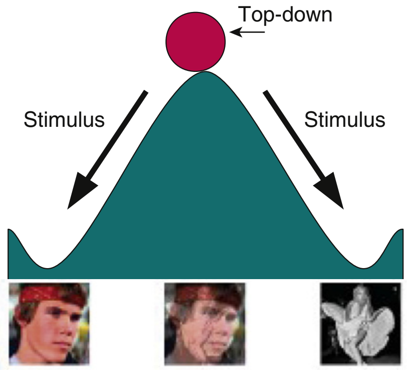


Figure 1. The closed-loop experimental design used by Cerf *et al.* [2], in which the stimulus image is continuously updated based on the current neural activity.



Current Biology

Figure 2. Illustration of effect of the positive feedback in the experimental loop.

A small initial top-down bias can push the brain-state (represented by the ball) towards a dominant stimulus by the changes in the stimulus.

Similarly, if the stimuli can be chosen or generated during testing, the space of stimuli can be constrained much more efficiently, even when our assumptions about the relevant stimulus parameters are weaker. Closed-loop stimulus selection has been used in testing neurons in high-level visual cortical area STSa of monkeys, where it allowed the testing of a large stimulus set [8]. Stimulus optimisation goes a step further by generating the stimuli on-line to either maximise the neural response [9,10] or to maximally reduce the uncertainty about the response properties [11–13]. Closed-loop stimulus (pixel) optimisation has been used to test complex cells in primary visual cortex [9] and this method has been adapted to optimising auditory stimuli [14]. Including a model of the neural pathway leading to the neuron under investigation can also make the search more effective [15].

Is the closed-loop design useful for mind-reading (brain-state reading)? Once we have a description of neural selectivity in the form of a conditional probability distribution of the responses given the stimulus or stimulus parameters, we can use Bayesian methods to decode or interpret any neural activity pattern in the form of a conditional distribution across the possible stimuli given the neural activity [16,17]. A visual reconstruction [18,19] may not be possible for abstract representations. The closed-loop design does not seem useful for such decoding, as it makes it harder to distinguish the effect of the stimulus and the brain state. An exciting potential application for closed-loop design lies in using it as a ‘semantic typewriter’, a thought-driven concept browser. Instead of choosing between two stimuli, it might be possible to decode neural activities corresponding to categories, and displaying exemplars or words, gradually zooming in on the precise meaning.

The combination of recording human single high-level neurons and the closed-loop design, illustrated by the new work of Cerf *et al.* [2], is a promising research direction, especially for exploring the uniquely human semantic, symbolic and linguistic representations of our brain.

Reference

1. Quian Quiroga, R., Reddy, L., Kreiman, G., Koch, C., and Fried, I. (2005). Invariant visual representation by single neurons in the human brain. Nature *435*, 1102–1107.
2. Cerf, M., Thiruvengadam, N., Mormann, F., Krasov, A., Quian Quiroga, R., and Koch, C. (2010). On-line, voluntary control of human temporal lobe neurons. Nature *467*, 1104–1108.
3. Reynolds, J.H., and Chelazzi, L. (2004). Attentional modulation of visual processing. Annu. Rev. Neurosci. *27*, 611–647.
4. Walther, D.B., and Koch, C. (2007). Attention in hierarchical models of object recognition. Progr. Brain Res. *165*, 57–78.
5. Ringach, D.L., Hawken, M.J., and Shapley, R. (2002). Receptive field structure of neurons in monkey primary visual cortex revealed by stimulation with natural image sequences. J. Vis. *2*, 12–24.
6. Schwartz, O., Pillow, J.W., Rust, N.C., and Simoncelli, E.P. (2006). Spike-triggered neural characterization. J. Vis. *6*, 484–507.
7. Wu, M.C.K., David, S.V., and Gallant, J.L. (2006). Complete functional characterization of sensory neurons by system identification. Annu. Rev. Neurosci. *29*, 477–505.
8. Földiák, P., Xiao, D.K., Keysers, C., Edwards, R., and Perrett, D.I. (2004). Rapid serial visual presentation for the determination of neural selectivity in area STSa. Prog. Brain Res. *144*, 107–116.
9. Földiák, P. (2001). Stimulus optimisation in primary visual cortex. Neurocomputing *38-40*, 1217–1222.
10. Edin, F., Machens, C., Schutze, H., and Herz, A. (2004). Searching for optimal sensory signals: Iterative stimulus reconstruction in closed-loop experiments. J. Comp. Neurosci. *17*, 47–56.
11. Sahani, M. (1997). Interactively exploring a neural code by active learning. Neural Information and Coding Workshop, Snowbird, Utah. http://bit.ly/fmdhFX
12. Benda, J., Gollisch, T., Machens, C.K., and Herz, A.V.M. (2007). From response to stimulus: adaptive sampling in sensory physiology. Curr. Opin. Neurobiol. *17*, 430–436.
13. Lewi, J., Butera, R., and Paninski, L. (2009). Sequential optimal design of neurophysiology experiments. Neural Comput. *21*, 619–687.
14. O’Connor, K.N., Petkov, C.I., and Sutter, M.L. (2005). Adaptive stimulus optimization for auditory cortical neurons. J. Neurophys. *94*, 4051–4067.
15. Lehky, S.R., Sejnowski, T.J., and Desimone, R. (1992). Predicting responses of nonlinear neurons in monkey striate cortex to complex patterns. J. Neurosci. *12*, 3568–3581.
16. Oram, M.W., Földiák, P., Perrett, D.I., and Sengpiel, F. (1998). The ‘Ideal Homunculus’: Decoding neural population signals. Trends Neurosci. *21*, 259–265.
17. Koyama, S., Chase, S.M., Whitford, A.S., Velliste, M., Schwartz, A.B., and Kass, R.E. (2010). Comparison of brain-computer interface decoding algorithms in open-loop and closed-loop control. J. Comp. Neurosci. *29*, 73–87.
18. Miyawaki, Y., Uchida, H., Yamashita, O., Sato, O., Morito, Y., Tanabe, H.C., Sadato, N., and Kamitani, Y. (2008). Visual image reconstruction from human brain activity using a combination of multiscale local image decoders. Neuron *60*, 915–929.
19. Naselaris, T., Prenger, R.J., Kay, K.N., Oliver, M., and Gallant, J.L. (2009). Bayesian reconstruction of natural images from human brain activity. Neuron *63*, 902–915.

School of Psychology, University of St
Andrews, St Andrews, KY16 9JP, UK.
E-mail: Peter.Foldiak@st-andrews.ac.uk

DOI: 10.1016/j.cub.2010.12.023