Journal Club

Editor's Note: These short reviews of recent *JNeurosci* articles, written exclusively by students or postdoctoral fellows, summarize the important findings of the paper and provide additional insight and commentary. If the authors of the highlighted article have written a response to the Journal Club, the response can be found by viewing the Journal Club at www.jneurosci.org. For more information on the format, review process, and purpose of Journal Club articles, please see http://jneurosci.org/content/preparing-manuscript#journalclub.

How Does the Brain Tell Self-Motion from Object Motion?

Benedict Wild

Cognitive Neuroscience Laboratory, German Primate Center and Göttingen Graduate School for Neurosciences, Biophysics, and Molecular Biosciences, Göttingen University, 37077 Göttingen, Germany

Review of Sasaki et al.

Imagine a football player who has to run across the field, evade opponent players, and eventually catch the ball that is passed to him by his teammate. As he does this, his retinal image is a hodgepodge of moving components. The position of background objects (e.g., yard lines, goal posts, the cheering crowd) changes relative to the runner as he runs across the field, creating a so-called "optic flow" pattern on his retina; the opposing team's players move toward him from different angles; and, most importantly, the football eventually appears in his field of vision. To navigate and act within this ever-changing environment, the player must be able to distinguish retinal motion resulting from his own movements from the motion caused by moving objects. But how can the relevant motion components be selected and confounding components be discarded, given that all the information the visual system receives is the jumble of movements across the retina?

The dorsal part of the medial superior temporal area in the macaque cortex (area MSTd) has long been known to play an important role in motion perception in general (Graziano et al., 1994) and self-motion perception in particular (Britten, 2008). Individual neurons in this part of the brain are typically tuned for the direction of a moving stimulus or for a particular heading direction, as measured with stimuli that simulate optic flow (e.g., a field of random dots that move away from a single point, which determines the direction of the simulated self-motion). Furthermore, the area contains multimodal neurons that integrate visual information with vestibular information (Duffy, 1998; Gu et al., 2008) for a joint representation of self-motion. Interestingly, some multimodal MSTd neurons have the same preferred direction for visual and vestibular input ("congruent cells"), whereas others prefer different or even opposing directions ("opposite cells") (Gu et al., 2008). During simultaneous visual and vestibular stimulation, the neuronal sensitivity for discriminating heading directions (calculated based on signal detection theory) (see Britten et al., 1992) is decreased in such opposite cells compared with either visual or vestibular stimulation alone. Congruent cells, on the other hand, show an increased sensitivity in this "bimodal condition" compared with either unimodal condition (Gu et al., 2008). This raises the question of what the purpose of opposite cells may be.

In a recent publication in *The Journal of Neuroscience*, Sasaki et al. (2017) suggested that opposite cells play an important role in parsing object motion and self-motion-components from the overall retinal image motion. To investigate this, they recorded the activity of individual neurons in area MSTd in 2 macaque monkeys who were placed in a virtual-reality setting. The monkeys were seated on a platform that could passively be moved in 3D space, thus providing vestibular stimulation. Simultaneously, visual optic flow stimuli were presented through a 3D field of stars that simulated translational self-motion in one of eight directions in the frontoparallel plane. Additionally, a cluster of nine spheres, defined by increased dot density (the "object"), moved in one of eight possible directions through the visual world on some of the trials (Sasaki et al., 2017, their Fig. 1).

Sasaki et al. (2017) found that the influence of object motion on heading tuning (and the influence of heading direction on object motion tuning) differed between cell types: for congruent cells (50% of recorded cells), heading tuning was more consistent in the bimodal condition (vestibular and visual stimulation) than in the visual-only condition, but object motion direction tuning was more consistent in the visual-only condition. Conversely, for opposite cells (~18% of recorded cells), heading tuning was stronger in the visual-only condition, whereas object-motion-direction tuning was more consistent in the bimodal condition. This makes intuitive sense: if a cell's preferred heading

direction is the same for visual and for vestibular stimuli (i.e., congruent cells), then adding vestibular information will help the cell to maintain its normal tuning curve in the face of visual input that is confounded by a moving object. For cells that have opposing preferences for visual and vestibular heading information (i.e., opposite cells), bimodal stimulation flattens the tuning curve, thereby decreasing the selectivity for heading direction. Furthermore, the preferred direction for moving objects typically is opposite to the preferred heading direction because self-motion in one direction (e.g., to the left) means that the retinal image, including individual objects, moves in the opposite direction (e.g., to the right). A cell that is tuned for heading to the left should therefore also be tuned for an object moving to the right. Thus, in opposite cells, vestibular heading tuning is aligned with object motion direction tuning, as both are the opposite of visual heading tuning.

How do these different cell types influence the way that heading direction and object motion direction are represented by a population of MSTd neurons? To determine which specific stimulus most likely elicited a given neural population response, the authors first computed the joint probability of a specific heading and a specific object motion given the population response. Despite some simplifying assumptions, this decoder was able to accurately estimate both the heading and the object motion. However, this strategy would become computationally expensive if there were several objects moving through the scene (which is often the case in real-life scenes), as this would require the brain to calculate a multidimensional probability distribution.

This problem can be solved, however, by a mathematical procedure called marginalization, which determines the probability of one event (e.g., a specific heading direction) independent of a second event (e.g., object motion) that modulates the probability of the first event. While models have previously suggested that this process can be implemented in the brain (Beck et al., 2011), Sasaki et al. (2017) considered an approximation of marginalization, which, they claimed, would be more intuitive in terms of neuronal implementation. To this end, they first tried to decode heading direction from the responses of bimodal neurons by calculating the likelihood of a given heading direction (or object motion direction) as the sum of each neuron's response, weighted by either its visual or its vestibular tuning curve (Jazayeri and Movshon, 2006). This approach was not successful, however, possibly because of the diversity of tuning properties across the neuronal population. Approximate linear marginalization (ALM) (Kim et al., 2016) differs from the traditional likelihood computation in that it uses a regression model to find the optimal weights with which each neuron influences the overall likelihood of a specific heading direction. Applying this procedure to either a subset of the recorded MSTd neurons (all opposite cells and an equal number of randomly selected congruent cells) or to the whole sample resulted in a much better decoding performance. The decoding error was also smaller in the bimodal condition than in the visual-only condition, showing that the algorithm uses vestibular information to improve its decoding accuracy. The profile of the decoding weights across the population was similar to the neurons' tuning curves, suggesting that the brain uses information it already has when determining how much each cell contributes to the population's representation of the stimulus. However, the ALM algorithm appeared to apply a gain factor to the neurons that was not inherent to the neurons' tuning properties.

In summary, Sasaki et al. (2017) showed that a new algorithm, which had previously been developed for a population of simulated MSTd neurons (Kim et al., 2016), can be applied to a population of real MSTd neurons to distinguish self-motion from object motion. These findings raise the question of whether such an algorithm can actually be implemented by the brain and, if so, how the brain could learn the weights that are assigned to each neuron. ALM learned the weights by being trained on 500 simulated trials for each stimulus condition, attempting to minimize the difference between the true probability distribution and the algorithm's estimation of this distribution. During development, the brain does not have direct access to the true probability distribution to quickly learn the correct decoding weights. Instead, it acts based on its own estimation and then has to infer correct and incorrect judgments based on feedback to the actions it took. This should in principle allow decision-making areas in the brain to learn how to optimally read out the population response of MSTd.

Another point in question is why the authors put so much emphasis on marginalization being implemented through a linear decoder, at the cost of being only an approximation. They state that, for the purpose of modeling, they assume the brain to be "limited to processing neural responses linearly" (Kim et al., 2016) and that nonlinear transformations "may be difficult for the brain to implement" (Sasaki et al., 2017). However, there is evidence that marginalization can be implemented in neural circuits through divisive normalization (Beck et al., 2011), a widespread nonlinear computation where neuronal responses are inhibited, and thus effectively normalized, by the summed activity of a pool of neurons (for review, see Carandini and Heeger, 2011). Furthermore, area MST has been suggested to integrate its visual input from the middle temporal area in a nonlinear manner (Mineault et al., 2012), raising additional doubts as to why decoding of the output of MST would have to be a linear approximation. Thus, although ALM can decode heading and object motion from a population response in area MSTd, this does not guarantee that the brain actually implements this specific computation.

What additional strategies could the brain use to distinguish self-motion from object motion? In many cases, self-motion is caused by actions of the individual, such as walking or running, and these actions provide extraretinal information, such as the stimulation of proprioceptive sensors in moving body parts, or efference copies of motor commands to other parts of the brain. These can be used to compensate for the effects of self-motion on the retinal image flow (e.g., Crowell et al., 1998), so that any uncompensated retinal motion is likely due to moving objects (Wallach, 1987). But even when being moved passively (e.g., while riding a train and looking out of the window), eye movements can provide additional information about object motion during self-motion (Warren and Rushton, 2007). This is particularly interesting, as MSTd neurons also carry signals about eye movements (Newsome et al., 1988).

In conclusion, Sasaki et al. (2017) provided evidence that area MSTd can help our football player achieve his goal. Information that is encoded by MSTd neurons can be used to compute the player's own movement running across the field, even as his perceived optic flow patterns are disrupted by the movements of other players and the football. Similarly, MSTd can represent the ball's trajectory, even though its motion on the retina is distorted by the player's own movements. This can be achieved by a neuronal population of congruent and opposite multisensory cells, whose responses can be decoded in a way that approximates the

mathematical process of marginalization to accurately estimate a single variable of interest. Whether this is the computation actually performed by the brain needs to be investigated in more detail. Showing that decision-related brain areas higher up in the cortical processing hierarchy represent heading or object motion information in a manner that is consistent with the read-out predicted by ALM would provide evidence in favor of this hypothesis.

References

Beck JM, Latham PE, Pouget A (2011) Marginalization in neural circuits with divisive normalization. J Neurosci 31:15310–15319. CrossRef Medline

Britten KH (2008) Mechanisms of self-motion perception. Annu Rev Neurosci 31:389–410. CrossRef Medline

Britten KH, Shadlen MN, Newsome WT, Movshon JA (1992) The analysis of visual motion: a comparison of neuronal and psychophysical performance. J Neurosci 12:4745–4765. Medline

Carandini M, Heeger DJ (2011) Normalization as a canonical neural computation. Nat Rev Neurosci 13:51–62. CrossRef Medline

Crowell JA, Banks MS, Shenoy KV, Andersen RA (1998) Visual self-motion perception during head turns. Nat Neurosci 1:732–737. CrossRef Medline

Duffy CJ (1998) MST neurons respond to optic flow and translational movement. J Neurophysiol 80:1816–1827. CrossRef Medline

Graziano MS, Andersen RA, Snowden RJ (1994) Tuning of MST neurons to spiral motions. J Neurosci 14:54–67. Medline

Gu Y, Angelaki DE, DeAngelis GC (2008) Neural correlates of multisensory cue integration in macaque MSTd. Nat Neurosci 11:1201–1210. CrossRef Medline

Jazayeri M, Movshon JA (2006) Optimal representation of sensory information by neural populations. Nat Neurosci 9:690–696. CrossRef Medline

Kim HR, Pitkow X, Angelaki DE, DeAngelis GC (2016) A simple approach to ignoring irrelevant variables by population decoding based on multisensory neurons. J Neurophysiol 116:1449–1467. CrossRef Medline

Mineault PJ, Khawaja FA, Butts DA, Pack CC (2012) Hierarchical processing of complex motion along the primate dorsal visual pathway. Proc Natl Acad Sci U S A 109:E972–E980. CrossRef Medline

Newsome WT, Wurtz RH, Komatsu H (1988) Relation of cortical areas MT and MST to pursuit eye movements: II. Differentiation of retinal from extraretinal inputs. J Neurophysiol 60:604–620. CrossRef Medline

Sasaki R, Angelaki DE, DeAngelis GC (2017) Dissociation of self-motion and object motion by linear population decoding that approximates marginalization. J Neurosci 37:11204–11219. CrossRef Medline

Wallach H (1987) Perceiving a stable environment when one moves. Annu Rev Psychol 38:1–27. CrossRef Medline

Warren PA, Rushton SK (2007) Perception of object trajectory: parsing retinal motion into self and object movement components. J Vis 7:2.1–11. CrossRef Medline