Trends in Cognitive Sciences

CellPress
REVIEWS

Forum

Dynamic Neural Representations: An Inferential Challenge for fMRI

Avniel Singh Ghuman[1,*] and Alex Martin[2]

Measures of brain activity with high temporal resolution have shown that the information represented in a single brain region undergoes dynamic changes on the scale of milliseconds. This dynamic process presents a unique inferential challenge to low temporal resolution neural measures, such as BOLD fMRI. Potential solutions for fMRI requiring further investigation and development are discussed.

Brain connectivity is highly recurrent at all levels. Increasing evidence obtained with modalities that can record neural activity on the scale of milliseconds with high spatial resolution [e.g., single unit recordings and intracranial electroencephalography (iEEG)] suggests that a functional consequence of highly recurrent anatomical connectivity is that representations can change over time within a single patch of cortex. For example, a recent study showed that while the initial burst of activity in V1 neurons have responses consistent with their classic receptive fields, activity 50–100 ms later shows the emergence of extraclassical effects such as contour integration, likely as a consequence of feedback from V4 [1]. Furthermore, single unit activity in monkey temporal cortex transitions in time from initially only being sensitive to whether a stimulus is a face or not, to also becoming sensitive to the particular face being shown approximately 50 ms later [2]. These dynamic changes lead to unique challenges for low temporal resolution measures of brain activity, such as provided by blood-oxygen-level dependent (BOLD) fMRI using typical parameters, which rely on the assumption of a stable representation over these time windows. Specifically, as it is typically implemented (Box 1), fMRI smears over these dynamics and, as a result, is more or less sensitive to time-dependent, qualitative differences in representation. This differential sensitivity to temporal dynamics can, in turn, lead to inferential issues regarding the representational, or computational role, of a particular patch of cortex.

Recent examples highlight the interpretational issues that can arise when using low temporal resolution measures of brain activity for evaluating the information represented in a specific region of the brain. For example, a recent MRI adaptation study found evidence in support of the idea that the visual word form area (VWFA) contained whole word templates that allowed for equivalent differentiation of words that are only one letter apart, as well as for words that are completely different [3]. In contrast, a recent fMRI decoding study [multivoxel pattern analysis (MVPA)] found a graded effect in which differentiation of the VWFA signal depended on the orthographic similarity of individual words [4]. Thus, whereas the adaptation study suggested that VWFA word processing is governed by whole word templates, the MVPA results suggested that VWFA processing is dependent on an organization by orthographic similarity.

A recent iEEG study provides a resolution to these conflicting fMRI findings by showing that the characteristics of word processing in the VWFA changes over short time windows [5]. Specifically, from approximately100–200 ms, individual words that are orthographically completely different can be distinguished from one another, but words that are only one letter apart cannot, consistent with an organization by orthographic similarity. In a later stage of processing, however, from about 250–500 ms, words that are one letter apart and words that are completely different can be discriminated from one another to a similar degree, consistent with an organization based on whole word templates (Figure 1A). These findings suggest that the two previously discussed fMRI studies may have been differentially sensitive to earlier representations, organized by orthographic similarity, versus later representations, organized like whole word templates, for reasons yet to be determined in this particular case, though possibly due to different experimental designs. The dynamic change in representation revealed by iEEG leads to a model of the role of VWFA in word processing in which the initial representation is organized by orthographic similarity and becomes fully individuated at a later time, perhaps as a result of recurrent interactions and constraints from downstream regions that process other aspects of word information, such as phonology and semantics.

Studies of the fusiform face area (FFA) provide another example of the interpretational issues that can arise when neural activity is measured using an imaging modality with relatively poor temporal resolution, thereby potentially masking different aspects of face representation. Recent evidence from iEEG recordings of FFA activity [6] show that during an early 100–250 ms time window, the general category of ‘faces’ could be clearly distinguished from other object categories (such as houses, hammers, and bodies), whereas individual faces could not be distinguished from one another. This situation changed during the next 250 ms window when individual faces could be distinguished (Figure 1B). This suggests that individual face information in the FFA is a result of recurrent processing between this region and other parts of the face processing network.

Box 1. Temporal Resolution of fMRI

It is generally assumed that fMRI is unable to detect rapid changes in neural representations due to the sluggishness of the hemodynamic response function (HRF). However, the ability to detect events that are close in time is not dependent on the 'speed' of the HRF, but rather the temporal variability of the estimated HRF. Over space (across voxels and regions) the variability is on the order of seconds, which is problematic for determining the temporal order of neural events in one region versus another. However, if the HRF is well behaved when looking at a single region, with variability on the order of 10s or 100s of milliseconds including measurement error, then potentially dynamic shifts in the representation in a region could be accurately measured.

Using typical scan parameters the temporal variance of the HRF estimate, due to both measurement and hemodynamic variability, is approximately 2 s [12]. With that much variance, it is impractical to assess the temporal order of events that are, for example, separated by 100 ms. One potential solution to this problem is to focus on a single region of interest and modulate the task timing in conjunction with using much faster scanning parameters. Recently, temporal variance of 200 ms was reported in primary visual cortex using a 100 ms repetition time (TR) in a rapid event-related design [9], in which case detecting the difference between events separated by 100 ms would be feasible using a practical number of trials. It is unclear if these results will hold in other regions because large variability in HRF dynamics across the brain. If the HRF variability in other regions is determined and is relatively small, and other challenges associated with scanning with very short TRs overcome, eventually it may become feasible to detect dynamic changes in representations within a single region using fMRI.

In contrast, studies using fMRI have provided conflicting results regarding the specificity of the information represented in FFA, perhaps due to mixing of neural activity across the processing dynamics illustrated by the iEEG study described above. Early fMRI studies examining exemplar-level sensitivity found that the FFA was highly selective for faces relative to other categories of object, but did not distinguish between individual faces [7]. This and related findings lead to the suggestion that the FFA functions as a general face detector, discriminating faces from other objects, but not one face from another. Subsequent studies, however, have provided mixed results, with some studies providing evidence consistent with this idea, while others have found evidence for individual face classification [8]. While a consensus is starting to converge towards the FFA having a role in recognizing individual faces, these mixed results highlight the difficulty that fMRI-based measurements have in finding exemplar-level information when the representation changes over time. Models based on fMRI alone either neglect the role of FFA in individual face representation, reflecting the null results, or suggest that neural populations in the FFA can code for individual faces, unaware of the possibility that

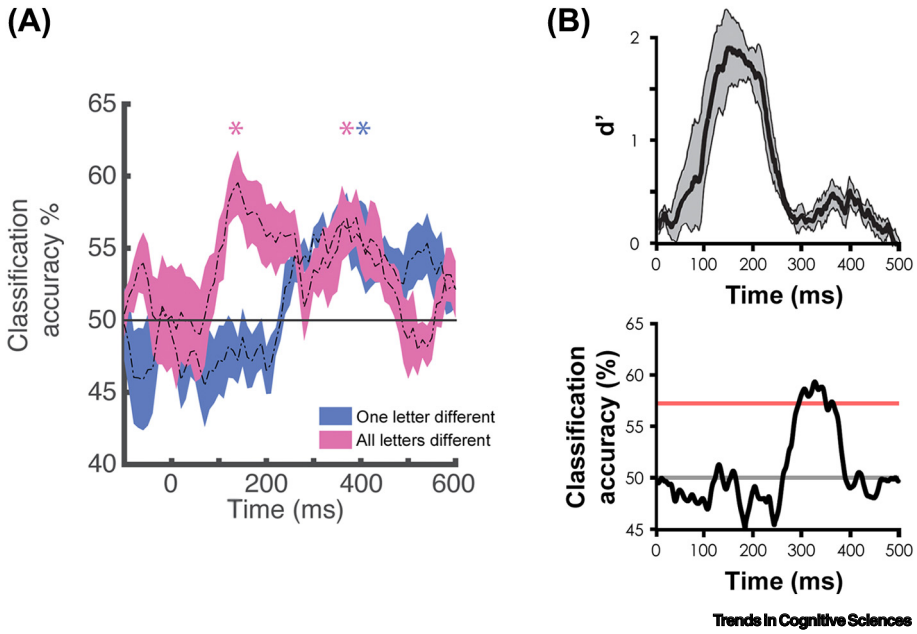

Figure 1. Dynamic Word and Face Representations in the Fusiform. Temporal dynamics of word decoding in the visual word form area (A) and face decoding from the fusiform face area (B) using intracranial electroencephalography in humans. (A) Example time courses from one subject with an electrode placed on the visual word form area [5]. At this electrode the information available at an early time window (100–200 ms) allowed for distinguishing only words that were completely different orthographically (e.g., hint versus dome; pink time course). At a later time (250–500 ms) words that differ only by one letter (e.g., hint versus lint; blue time course), as well as words that were completely different, could be decoded. (B) Time courses showing category-level decoding of faces (top) and exemplar-level decoding of facial identity (bottom) from an example subject with an electrode placed on the fusiform face area [6]. At this electrode the information available at an early time window (100–250 ms) allowed for distinguishing faces from other categories of objects, but not from one another. Individual faces could, however, be distinguished at a later time window (250–500 ms).

this code may only emerge as a late consequence of recurrent interactions with other parts of the face processing network.

Evidence from modalities with millisecond resolution, such as iEEG, magnetoencephalography, and single unit recordings, provide strong evidence that the nature of the information represented in a specific region of cortex, and the way that information is processed, can change over very short periods of time. These findings thus provide a clear challenge to claims about the nature of information or process represented in a specific cortical region based solely on relatively sluggish modalities like BOLD fMRI using typical scan parameters. Resolving this challenge will be critical if fMRI is to be used to understand and constrain computational models of the brain. Advances in fMRI acquisition may allow for a finer parsing of the temporal evolution of the hemodynamic response [9,10], which may eventually allow for the separation of the representations seen during earlier and later processing windows (Box 1). New innovations in laminar-level imaging may also be helpful (e.g., [11]). For example, within a specific region such as the FFA or VWFA, earlier representations may be more likely to reflect bottom-up processing, potentially relying on projections to different layers of the cortex than later representations, which, in turn, may more likely reflect the output of top-down and recurrent processes. fMRI measures that can separate out the response in different cortical layers could therefore help to elucidate the processes that lead to the emergence of qualitatively different types of representations within a single brain region.

Acknowledgments

We thank Peter Bandettini for valuable discussion and comments. We gratefully acknowledge the support of the National Institute of Mental Health under R01 MH107797 (to A.G.) and ZIA MH002920–09 (to A.M.) and National Science Foundation under 1734907 (to A.G.).

[1]Department of Neurological Surgery and the Center for the Neural Basis of Cognition, University of Pittsburgh, Pittsburgh, PA, USA
[2]Section on Cognitive Neuropsychology, Laboratory of Brain and Cognition, NIMH, NIH, Bethesda, MD, USA

*Correspondence:
ghumana@upmc.edu (A.S. Ghuman).

https://doi.org/10.1016/j.tics.2019.04.004

References

1. Liang, H. *et al.* (2017) Interactions between feedback and lateral connections in the primary visual cortex. *Proc. Natl. Acad. Sci. U. S. A.* 32, 8637–8642
2. Sugase, Y. *et al.* (1999) Global and fine information coded by single neurons in the temporal visual cortex. *Nature* 400, 869–873
3. Glezer, L.S. *et al.* (2009) Evidence for highly selective neuronal tuning to whole words in the "visual word form area". *Neuron* 62, 199–204
4. Baeck, A. *et al.* (2015) Influence of lexical status and orthographic similarity on the multi-voxel response of the visual word form area. *Neuroimage* 111, 321–328
5. Hirshorn, E.A. *et al.* (2016) Decoding and disrupting left midfusiform gyrus activity during word reading. *Proc. Natl. Acad. Sci. U. S. A.* 113, 8162–8167
6. Ghuman, A.S. *et al.* (2014) Dynamic encoding of face information in the human fusiform gyrus. *Nat. Commun.* 5, 5672
7. Kriegeskorte, N. *et al.* (2007) Individual faces elicit distinct response patterns in human anterior temporal cortex. *Proc. Natl. Acad. Sci. U. S. A.* 104, 20600–20605
8. Kanwisher, N. (2017) The quest for the FFA and where it led. *J. Neurosci.* 37, 1056–1061
9. Lin, F.H. *et al.* (2018) Relative latency and temporal variability of hemodynamic responses at the human primary visual cortex. *NeuroImage* 164, 194–201
10. Misaki, M. *et al.* (2013) Accurate decoding of sub-TR timing differences in stimulations of sub-voxel regions from multi-voxel response patterns. *NeuroImage* 66, 623–633
11. Huber, L. *et al.* (2017) High-resolution CBV-fMRI allows mapping of laminar activity and connectivity of cortical input and output in human M1. *Neuron* 96, 1253–1263
12. Saad, Z.S. *et al.* (2001) Analysis and use of FMRI response delays. *Hum. Brain Mapp.* 13, 74–93