

# The hemodynamic inverse problem: Making inferences about neural activity from measured MRI signals

Randy L. Buckner\*

Departments of Psychology, Radiology, and Anatomy and Neurobiology, Howard Hughes Medical Institute, Washington University, St. Louis, MO 63130

Cognitive events rely on rapid information flow among interdependent brain areas. Estimates from single-unit physiology in monkeys and human electrophysiological scalp recordings suggest communication between brain areas takes place on the order of tens to hundreds of milliseconds (1). A challenge to the study of human cognition has been to develop noninvasive methods to measure the temporal orchestration of information processing. Limitations were initially imposed by the imaging devices themselves. Positron emission tomography (PET) methods based on short-lived isotopes, for example, take about a minute to make a measurement (2). The advent of functional MRI (fMRI) changed the landscape of human brain imaging by providing a tool that can make measurements on the sub-second time scale (3, 4). However, the use of fMRI solves only part of the problem. An even greater challenge surrounds the temporal sluggishness of the correlates of neuronal activity to which modern imaging devices are sensitive. Most fMRI studies measure brain activity indirectly through changes in blood vasculature that accompany neuronal activity. Although the exact origins of these vascular changes are debated (5, 6), they are temporally slow, begin seconds after a neuronal event, and last for tens of seconds (7). For isolated cognitive acts, which can often be completed in under a second, the sluggish nature of the measured response means that a signal is detected after the neuronal event has subsided. It is against this backdrop that the work of Bellgowan *et al.* in a recent issue of PNAS (8) can be appreciated. Their study uses a modeling approach to make inferences about neural activity timing differences of  $\approx 100$  milliseconds. To understand the basis of this surprisingly high temporal resolution, it is important to consider the evolution of fMRI.

When fMRI first appeared as a viable method for cognitive neuroscience (3, 4), there was already considerable enthusiasm that it could provide revolutionary spatial and temporal resolution. Surprisingly, with some notable exceptions (7), initial fMRI methods based their measurements on paradigms that collected minute-long epochs of averaged brain activity. There

were probably a number of reasons for this choice that include the historical precedent from PET, readily available data analysis strategies (9), and the power of such paradigms.

Among the first major advances of fMRI over earlier methods was the advent of event-related fMRI (refs. 10–12; for review, see ref. 13). Event-related methods allow individual cognitive events to be isolated by recording the MRI signal in relation to the onsets of isolated trial types. To circumvent the

## Hemodynamic response shows a roughly linear relation to underlying neuronal activity.

sluggish vascular response, often referred to as the hemodynamic response, trials were initially spaced widely apart. More sophisticated procedures that allowed rapidly presented trials soon followed. These procedures used various analytic methods to estimate the hemodynamic response within the context of sequential, overlapping responses (14, 15). By using such paradigms, fMRI exploration of cognitive function now possessed the temporal resolution to mix different event types together and sort events based on subject responses such as whether an error was made (16) or an event was remembered (17, 18).

A contemporary and related advance came with an understanding of how the hemodynamic response summates in time (19) and over additive events (14). In seminal work, Geoffrey Boynton, David Heeger, and colleagues (19) used manipulation of the contrast and timing of visual stimuli to show that the hemodynamic response extended in time proportionate to the duration of neural activity and also increased in amplitude proportionate to the change in intensity of neural activity (as inferred by manipulations of visual stimuli; see also ref. 20). In their studies, while subtle departures from a perfect correspondence were observed, the measured hemodynamic response behaved in a manner

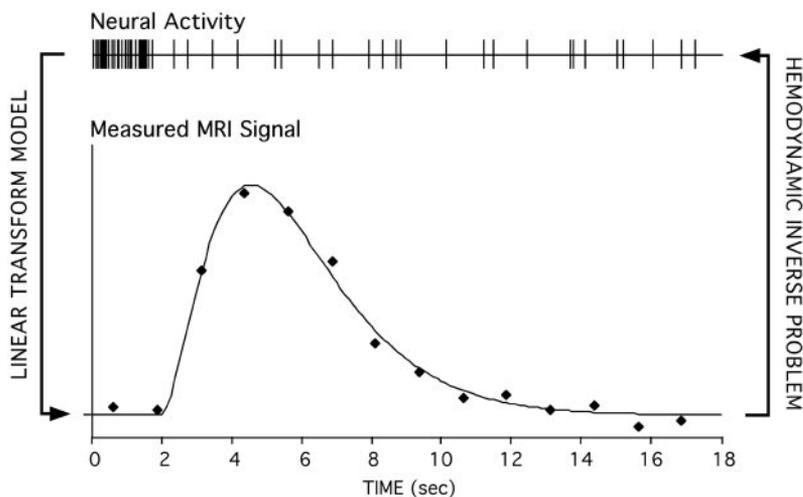
that suggested an approximately linear relation to underlying neuronal activity (for review, see ref. 21). In making these observations, a relatively simple modeling procedure was also provided to estimate parameters of the hemodynamic response based on a gamma function. The use of a linear transform model to make inferences about neural activity from hemodynamic measurements serves as a foundation for the present work by Bellgowan *et al.* (8).

The finding that the hemodynamic response shows a roughly linear relation to underlying neuronal activity has important consequences for what inferences can be drawn from the measured hemodynamic signal. Most importantly, it sets the stage for what one might call the “hemodynamic inverse problem.” The hemodynamic inverse problem refers to the challenge of making valid and precise estimates of underlying neural activity from the measured hemodynamic response (see Fig. 1). The demonstration of a forward relation between the stimulus, neuronal activity, and hemodynamic response (19) suggested that, within the context of event-related designs, it might be possible to infer the amplitude and timing of the underlying neural activity. For example, hemodynamic timing delays may represent temporally shifted, but proportionate, timing delays in underlying neuronal activity.

Bellgowan *et al.* (8) present a study representing progress in solving the hemodynamic inverse problem. In their study, subjects were imaged with event-related fMRI while deciding whether letter strings represented words (e.g., “ketchup”) or nonwords (e.g., “techper”). Referred to as the lexical decision task, this task relies on multiple language processing routes and decision processes that vary as a function of factors such as whether the string is a word or a nonword. A second manipulation was also included in their study: letter strings were rotated. Some letter strings were presented in standard horizontal format, whereas others were rotated by as

See companion article on page 1415 in issue 3 of volume 100.

\*To whom correspondence should be addressed at: Campus Box 1125, Department of Psychology, Howard Hughes Medical Institute, Washington University, One Brookings Drive, St. Louis, MO 63130. E-mail: rbuckner@artsci.wustl.edu.



**Fig. 1.** A schematic illustrates the linear transform model that specifies the relation between averaged neural activity and the temporally prolonged hemodynamic response measured by fMRI. Many properties of the relation between the two are still being explored, such as whether the MRI signal reflects averaged neural spiking or synaptic events (e.g., refs. 6, 21, and 28). The hemodynamic inverse problem concerns inferences in the opposite direction, exploring whether the measured hemodynamic response can be used to estimate underlying neural events. In the paper by Bellgowan *et al.* in a recent issue of PNAS (8), a model fit to the data was used to estimate the onset and width of the hemodynamic response and to make inferences about the timing and duration of underlying neural activity.

much as 120°. This second manipulation caused variation in perceptual processes. Of importance, extending from earlier studies of visual and motor regions (for examples, see refs. 22–24), Bellgowan *et al.* (8) estimated hemodynamic timing differences to make inferences about brain regions involved in language and decision-stage processing.

Two separate technical issues were overcome in making these estimates. The first issue surrounds variation in hemodynamic response timing across the brain. Although hemodynamic response timing reflects the timing of neuronal activity, a large confounding contribution comes from baseline timing differences that originate from the local vasculature's architecture (25, 26). These timing differences can be several seconds in magnitude, thus effectively swamping smaller temporal changes reflective of information processing. To account directly for baseline timing differences, maps of relative change in timing from one condition (e.g., non-words) to the next (e.g., words) were used. By using relative change as a measure, the temporally lagged and spatially variable hemodynamic response was normalized. The second issue surrounds the challenge of modeling the measured hemodynamic response to provide meaningful estimates. By using a modeling approach based on convolution of two separate functions,

they were able to make unbiased estimates of the timing delay as well as duration (width) of the hemodynamic response.

Results suggested that the hemodynamic timing estimates of certain regions tracked the task variables. For example, regions within prefrontal cortex, along the inferior frontal gyrus, showed an onset delay proportionate to the rotation of the letter strings and also an extended duration during nonword decisions, as compared with word decisions (see also ref. 27). Although interpretation of these differences is speculative, the results are consistent with a role of frontal cortex in decision processes that are extended in time. Of importance, in making these observations, they demonstrate hemodynamic timing differences of a few hundred milliseconds, a temporal resolution that seemed implausible when MRI-based hemodynamic imaging methods became available. Open issues remain, such as how to interpret width as contrasted to amplitude differences in the hemodynamic response and, more generally, what inferences can be drawn about the underlying neural activity from these measures.

Has fMRI based on measurement of hemodynamic response reached its limit? The achievement of Bellgowan *et al.* of subsecond temporal resolution in a cognitive paradigm is remarkable, given the limitations of the sluggish sig-

nal being measured. However, the temporal resolution presented still falls short of what might be needed to observe the complex spatial-temporal orchestration of brain activity that underlies cognition. Placing the accomplishment in context, analysis of activity between brain areas will likely require temporal resolution of 10 msec or less. Certain questions, including many that target mechanisms of information flow, focus their level of analysis on adjacent columns within areas and even between layers of cortex within columns. Absolute limitations of the methodology will be due, in part, to technical limitations of the imaging devices and artifacts encountered. For example, human brains pulsate, causing temporal blurring. Other barriers will likely surround the properties of the indirect hemodynamic signals that are being measured. Growing evidence suggests that events in the axon terminals and dendrites of neurons may underlie the signals observed with hemodynamic MRI measures (ref. 28; for review, see ref. 6). If this suggestion is correct, hemodynamic measures will be amenable to making inferences about net properties of components of neuronal events within an area, paralleling the local field potential (LFP) recorded by microelectrode studies and not the spiking activity itself.

For these reasons, it seems unlikely that noninvasive hemodynamic measures in humans will soon provide a means to visualize the millisecond timing of neuronal events. Other kinds of method, which can be used in combination with hemodynamic measures, will likely push the limits of temporal resolution in human measurement (for example, see refs. 29–31). Animal models that allow single-units as well as LFPs to be measured will be essential to future progress (for example, see ref. 28). Nonetheless, it is worthwhile to put skepticism in its place. The commentary that accompanied the first report of event-related fMRI (32) also contained concerns about whether the limits of MRI methods had been reached. Having surveyed the promise of the new methods being reported, which were based on averages over individual events, a note was made that “the imaging of the neural correlates of single and discrete mental events, such as *one* image or *one* word, remains a most desirable dream.” Within a year following the printing of the commentary, Wolfgang Richter, Seong-Gi Kim, and colleagues reported an analysis of isolated mental events (23).

1. Robinson, D. L. & Rugg, M. D. (1988) *Biol. Psychol.* **26**, 111–116.
2. Raichle, M. E. (1987) in *Handbook of Physiology, The Nervous System*, eds. Mountcastle, V. B. &

- Plum, F. (Oxford Univ. Press, New York), Vol. 5, pp. 643–674.
3. Kwong, K. K., Belliveau, J. W., Chesler, D. A., Goldberg, I. E., Weisskoff, R. M., Poncelet, B. P.,

- Kennedy, D. N., Hoppel, B. E., Cohen, M. S., Turner, R., *et al.* (1992) *Proc. Natl. Acad. Sci. USA* **89**, 5675–5679.
4. Ogawa, S., Lee, T. M., Kay, A. R. & Tank,

- D. W. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 9868–9872.
5. Raichle, M. E. (1998) *Proc. Natl. Acad. Sci. USA* **95**, 765–772.
  6. Logothetis, N. (2003) *J. Neurosci.*, in press.
  7. Blamire, A. M., Ogawa, S., Ugurbil, K., Rothman, D., McCarthy, G., Ellerman, J. M., Hyder, F., Rattner, Z. & Shulman, R. G. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 11069–11073.
  8. Bellgowan, P. S. F., Saad, Z. S. & Bandettini, P. A. (2003) *Proc. Natl. Acad. Sci. USA* **100**, 1415–1419.
  9. Bandettini, P. A., Jesmanowicz, A., Wong, E. C. & Hyde, J. S. (1993) *Magn. Reson. Med.* **30**, 161–173.
  10. Buckner, R. L., Bandettini, P. A., O'Craven, K. M., Savoy, R. L., Petersen, S. E., Raichle, M. E. & Rosen, B. R. (1996) *Proc. Natl. Acad. Sci. USA* **93**, 14878–14883.
  11. McCarthy, G., Luby, M., Gore, J. & Goldman-Rakic, P. (1997) *J. Neurophysiol.* **77**, 1630–1634.
  12. Friston, K. J., Fletcher, P., Josephs, O., Holmes, A., Rugg, M. D. & Turner, R. (1998) *NeuroImage* **7**, 30–40.
  13. Rosen, B. R., Buckner, R. L. & Dale, A. M. (1998) *Proc. Natl. Acad. Sci. USA* **95**, 773–780.
  14. Dale, A. M. & Buckner, R. L. (1997) *Hum. Brain Mapp.* **5**, 329–340.
  15. Clark, V. P., Maisog, J. M. & Haxby, J. V. (1998) *J. Neurophysiol.* **79**, 3257–3268.
  16. Carter, C. S., Braver, T. S., Barch, D. M., Botvinick, M. M., Noll, D. & Cohen, J. D. (1998) *Science* **280**, 747–749.
  17. Wagner, A. D., Schacter, D. L., Rotte, M., Koutstaal, W., Maril, A., Dale, A. M., Rosen, B. R. & Buckner, R. L. (1998) *Science* **281**, 1188–1191.
  18. Brewer, J. B., Zhao, Z., Desmond, J. E., Glover, G. H. & Gabrieli, J. D. (1998) *Science* **281**, 1185–1187.
  19. Boynton, G. M., Engel, S. A., Glover, G. H. & Heeger, D. J. (1996) *J. Neurosci.* **16**, 4207–4221.
  20. Heeger, D. J., Huk, A. C., Geisler, W. S. & Albrecht, D. G. (2000) *Nat. Neurosci.* **3**, 631–633.
  21. Heeger, D. J. & Ress, D. (2002) *Nat. Rev. Neurosci.* **3**, 142–151.
  22. Menon, R. S., Luknowsky, D. C. & Gati, J. S. (1998) *Proc. Natl. Acad. Sci. USA* **95**, 10902–10907.
  23. Richter, W., Ugurbil, K., Georgopolous, A. & Kim, S.-G. (1997) *NeuroReport* **8**, 3697–3702.
  24. Miezin, F. M., Maccotta, L., Ollinger, J. M., Petersen, S. E. & Buckner, R. L. (2000) *NeuroImage* **11**, 735–759.
  25. Lee, A. T., Glover, G. H. & Meyer, C. H. (1995) *Magn. Reson. Med.* **33**, 745–754.
  26. Robson, M. D., Dorosz, J. L. & Gore, J. C. (1998) *NeuroImage* **7**, 185–198.
  27. Henson, R. N., Price, C. J., Rugg, M. D., Turner, R. & Friston, K. J. (2002) *NeuroImage* **15**, 83–97.
  28. Logothetis, N. K., Pauls, J., Augath, M., Trinath, T. & Oeltermann, A. (2001) *Nature* **412**, 150–157.
  29. Heinze, H. J., Mangun, G. R., Burchert, W., Hinrichs, H., Scholz, M., Munte, T. F., Gos, A., Scherg, M., Johannes, S., Hundeshagen, H., et al. (1994) *Nature* **372**, 543–546.
  30. Snyder, A. Z., Abdullaev, Y. G., Posner, M. I. & Raichle, M. E. (1995) *Proc. Natl. Acad. Sci. USA* **92**, 1689–1693.
  31. Dale, A. M., Liu, A. K., Fischl, B. R., Buckner, R. L., Belliveau, J. W., Lewine, J. D. & Halgren, E. (2000) *Neuron* **26**, 55–67.
  32. Grabowski, T. J. & Damasio, A. R. (1996) *Proc. Natl. Acad. Sci. USA* **93**, 14302–14303.