# What We Can and What We Can't Do with fMRI

Nikos K. Logothetis, PhD[1,2]

[1]Max Planck Institute for Biological Cybernetics
Tübingen, Germany

[2]Imaging Science and Biomedical Engineering
University of Manchester
Manchester, United Kingdom

# Introduction

Functional activation of the brain can be detected with magnetic resonance imaging (MRI) by directly measuring tissue perfusion, blood-volume changes, or changes in the concentration of oxygen. The latter blood oxygenation level–dependent (BOLD) contrast mechanism (Logothetis, 2003; Logothetis and Wandell, 2004; Logothetis, 2008) is currently the mainstay of human neuroimaging. The interpretation of fMRI signals in brain research, and by extension, the utility of fMRI, critically depends on factors such as signal specificity and spatial and temporal resolution. Signal specificity ensures that the generated maps reflect actual neural changes, whereas spatial and temporal resolution determine our ability to discern the elementary units of the activated networks and the time course of various neural events.

Spatial specificity increases with increasing magnetic field and, for a given magnetic field, can be optimized by using pulse sequences that are less sensitive to signals from within and around large vessels (Fig. 1).

Spatiotemporal resolution is likely to keep increasing owing to several factors: the optimization of pulse sequences, the improvement of resonators, the use of higher magnetic fields, and the invention of intelligent strategies for parallel imaging (Logothetis, 2008).

fMRI may soon provide us with images of a fraction of a millimeter (e.g., 300 × 300 $\mu m^2$ with slice thickness of a couple of millimeters) on a routine basis, which amounts to ~2–3 orders of magnitude smaller voxel-volumes than those currently used in human imaging (Logothetis, 2008). With an increasing number of acquisition channels, such resolution may ultimately be attained in whole-head imaging protocols, yielding unparalleled maps of distributed brain activity in great regional detail and with reasonable temporal resolution of a couple of seconds. But would that be enough? The answer obviously depends on the scientific question and the spatial scale at which this question could be addressed.

To understand the functioning of the microcircuits in cortical columns, or of the cell assemblies in the
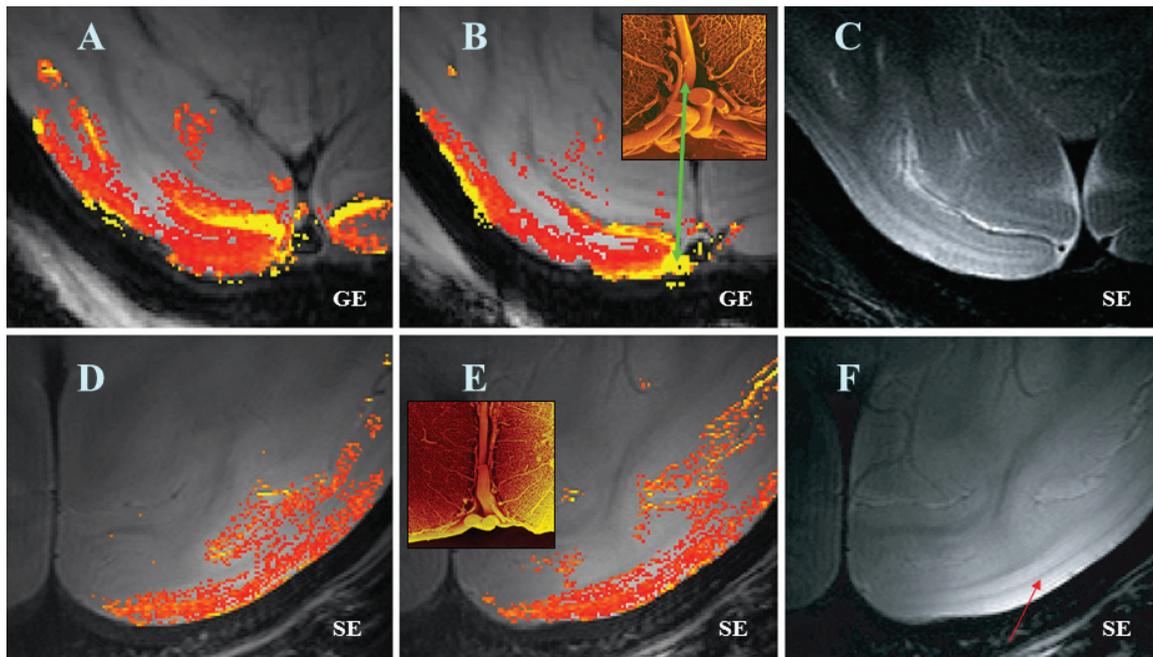


**Figure 1.** Specificity of GE-EPI and SE-EPI and examples of high-resolution GE-EPI and SE-EPI. **A**, and **B**: Two slices of GE-EPI demonstrating the high functional SNR of the images but also the strong contribution of macrovessels. The yellow areas (green arrows) are pial vessels, an example of which is shown in the inset with the SEM image. In-plane resolution, 333 × 333 $\mu m^2$; slice thickness, 2 mm. **C**, Anatomical scan, SE-EPI, 250 × 188 $\mu m^2$, 2 mm slice, with TE/TR = 70/3,000 ms. **D** and **E**, Two slices of SE-EPI showing a reduction of the vascular contribution at the pial side of the cortex. In-plane resolution, 250 × 175 $\mu m^2$; slice thickness, 2 mm. **F**, The anatomical scan is the SE-EPI used for obtaining the functional scans (TE/TR = 48/2,000 ms) but at different gray scale and contrast. The resolution of the anatomical scan permits the clear visualization of the Gennari line, the characteristic striation of the primary visual cortex. GE-EPI, gradient-echo echo-planar imaging; SE-EPI, spin-echo echo-planar imaging; TE/TR, echo time/repetition time. Logothetis et al. (2008), their Fig. 1, reprinted with permission. Copyright ©2008 Macmillan Publishers.

striosomes of basal ganglia, one must know a great deal about synapses, neurons, and their interconnections. In the same way, to understand the functioning of a distributed large-scale system, such as that underlying our memory or linguistic capacities, one must first understand the architectural units that organize neural populations of similar properties and how such units are interconnected.

With $10^{10}$ neurons and $10^{13}$ connections in the cortex alone, attempting to study dynamic interactions between subsystems at the level of single neurons would probably make little sense, even if it were technically feasible. Instead, it is probably much more important to better understand the differential activity of functional subunits—whether subcortical nuclei, or cortical columns, blobs, and laminae—and the instances of their joint or conditional activation. If so, whole-head imaging with a spatial resolution, say, of $0.7 \times 0.7$ mm$^2$ in slices of 1 mm thickness and a sampling time of a couple of seconds might prove optimal for addressing the vast majority of questions in basic and clinical research. This would be true even more because of the great sensitivity of the fMRI signal to neuromodulation. Neuromodulatory effects (e.g., those effected by arousal, attention, and memory) are slow and therefore lead to reduced spatiotemporal resolution and specificity (Motter, 1993; Luck et al., 1997).

## Activation Maps: What Do They Represent?

Does the activation of an area mean it is truly involved in the task at hand? This question implies that we understand what sort of neural activity in a given area would unequivocally show its participation in a studied behavior. But do we? It is usually alleged that cognitive capacities reflect the "local processing of inputs" or the "output" of a region, instantiated in the patterns of action potentials, with their characteristic frequency and timing. In principle, brain structures can be conceptualized as information-processing entities, with an input, a local-processing capacity, and an output. Yet, although such a scheme may describe the function of subcortical nuclei, its implementation in different areas of cortex is anything but straightforward.

We now know that the traditional cortical input–elaboration–output scheme (commonly presented as an instantiation of the tripartite perception–cognition–action model) is likely to be a misleading oversimplification (Logothetis, 2008). Research has shown that the subcortical input to the cortex is weak, the feedback is massive, the local connectivity

reveals strong excitatory and inhibitory recurrence, and the output reflects changes in the balance between excitation and inhibition rather than simple feedforward integration of subcortical inputs (Douglas and Martin, 2004). Therefore, when discussing the neural basis of hemodynamic signals, the properties of these excitation–inhibition networks (EINs) deserve special attention, and are discussed in "Modules and Their Microcircuits," below.

## Feedforward and Feedback Cortical Processing

Brain connectivity is mostly bidirectional. To the extent that different brain regions can be thought of as hierarchically organized processing steps, connections are often described as feedforward and feedback, forward and backward, ascending–descending, or alternatively, bottom-up and top-down. In the sensory systems, patterns of long-range cortical connectivity to some extent define feedforward and feedback pathways. The main thalamic input goes to the middle cortical layers, while second-order thalamic afferents as well as the nonspecific diffuse afferents from basal forebrain and brainstem are distributed diffusely and regionally, or over many cortical areas, respectively, and have their synapses mainly in superficial and/or deep layers. Cortical output has thalamic and other subcortical projections originating in layers VI and V, respectively, and corticocortical projections originate mostly from supragranular layers. The primary thalamic input innervates both excitatory and inhibitory neurons, and communication between all cell types includes horizontal and vertical connections within and between cortical layers. Such connections are divergent and convergent, so the final response of each neuron is determined by all feedforward, feedback, and modulatory synapses (Douglas and Martin, 2004).

Surprisingly, very few of the pyramid synapses are thalamocortical (less than 10–20% in the input layers of cortex, less than 5% across its entire depth, while in the primary visual cortex, thalamocortical synapses on stellate cells make up only ~5%). The remaining synapses originate from other cortical pyramidal cells. Pyramidal axon collaterals ascend back to and synapse in superficial layers, while others distribute excitation in the horizontal plane, forming a strongly recurrent excitatory network (Douglas and Martin, 2004).

The strong amplification of the input signal that this kind of positive feedback loop causes is tightly controlled by an inhibitory network interposed

among pyramidal cells and consisting of a variety of GABAergic interneurons. These interneurons can receive both excitatory and inhibitory synapses onto their somata and have only local connections. Approximately 85% of them, in turn, innervate the local pyramidal cells. Different GABA-ergic cells target different subdomains of neurons. Some, e.g., basket cells, which target somata and proximal dendrites, are excellent candidates for the role of gain-adjustment of the integrated synaptic response; others, e.g., chandelier cells, directly target the axons of nearby pyramidal neurons and appear to play a context-dependent role. That is, they can facilitate spiking during low-activity periods or act like gatekeepers that shunt most complex somatodendritic integrative processes during high-activity periods. Such nonlinearities might generate substantial dissociations between subthreshold population activity (and its concomitant metabolic demand) and the spiking of pyramidal cells.

## Modules and Their Microcircuits

Many structural, immunochemical, and physiological studies in all cortical areas examined so far suggest that the functional characteristics of a cortical module can be instantiated in a simple, basic EIN. This EIN is referred to as "canonical microcircuit" and has the following distinct features (Logothetis, 2008):

(1) The final response of each neuron is determined by all feedforward, feedback, and modulatory synapses;

(2) Transient excitatory responses may result from leading excitation (e.g., due to small synaptic delays or differences in signal-propagation speed), whereupon inhibition is rapidly engaged, followed by balanced activity;

(3) Net excitation or inhibition (E-I) might occur when the afferents drive the overall E-I balance in opposite directions; and

(4) Responses to large, sustained input changes may occur while maintaining a well-balanced E-I.

In the latter case, experimentally induced hyperpolarization of pyramidal cells may abolish their spiking without affecting the barrages of postsynaptic potentials. It is reasonable to assume that any similar hyperpolarization under normal conditions would decrease spiking of stimulus selective neurons without affecting presynaptic activity. In visual cortex, recurrent connections among spiny stellate cells in the input layers can provide a significant source of recurrent excitation. If driven by proportional E-I synaptic currents, the impact of

their sustained activity might, once again, minimally change the spiking of the pyramidal cells. This last property of microcircuits suggests that changes with balanced E-I are good candidates for mechanisms adjusting the overall excitability and the signal-to-noise ratio (SNR) of the cortical output. Thus, as for thalamic circuits, microcircuits (Guillery and Sherman, 2002; Sherman, 2005), depending on their mode of operation, can in principle act in two ways: as drivers, faithfully transmitting stimulus-related information, or as modulators, adjusting the overall sensitivity and context-specificity of the responses.

## E-I Networks and fMRI

The organization just discussed evidently complicates both the precise definition of the conditions that would justify assigning a functional role to an "active" area, and interpretation of the fMRI maps. Changes in E-I balance—whether they lead to net excitation, inhibition, or simple sensitivity-adjustment—inevitably and strongly affect regional metabolic energy demands and the concomitant regulation of cerebral blood flow (CBF). That is, such changes significantly alter the fMRI signal. A frequent explanation of fMRI data simply assumes an increase in the spiking of many task-specific or stimulus-specific neurons. This interpretation might be correct in some cases. However, the BOLD signal may also be increased as a result of balanced, proportional increases in the E-I conductances, potential concomitant increases in spontaneous spiking, but still without a net excitatory activity in stimulus-related cortical output. In the same manner, an increase in recurrent inhibition with concomitant decreases in excitation may reduce an area's net spiking output, but would it decrease its fMRI signal? Whether it does or not appears to depend on the brain region that is inhibited as well as on experimental conditions.

Direct hemodynamic measurements using positron emission tomography (PET) suggest that metabolism increases along with increased inhibition (Jueptner and Weiller, 1995, 1998). An exquisite example is the inhibition-induced increase in metabolism in the cat lateral superior olive (LSO) (Nudo and Masterton, 1986). Presynaptic activity in LSO is sufficient to show strong activations despite the ensuing spiking reduction. Also, similar increases in metabolism during the reduction in spike rates were observed during long-lasting microstimulation of the fornix, which induces sustained suppression of pyramidal cell firing in the hippocampus (Ackermann et al., 1984).

In contrast, human fMRI studies have reported hemodynamic and metabolic downregulation

accompanying neuronal inhibition in motor and visual cortices. These results suggest that the sustained negative BOLD response (NBR) is a marker of neuronal deactivation (Logothetis, 2002; Logothetis and Pfeuffer, 2004; Logothetis and Wandell, 2004; Shmuel et al., 2006). Similarly, experiments combining fMRI and electrophysiological experiments have revealed a clear correspondence between NBR and decreased population spiking in hemodynamically "negative" areas in the monkey primary visual cortex (Shmuel et al., 2006). The diversity of hemodynamic responses to neural inhibition obtained in different types of experiments is hardly surprising. It is primarily the result of the fact that regional inhibition itself might have a number of different causes, including:

(1) Early shunting of the weak cortical input, leading to a reduction of recurrent excitation rather than an increase in summed inhibition;

(2) Increased synaptic inhibition;

(3) Shunting of the cortical output through the axono–axonic connections of the Chandelier cells; or

(4) Any combination thereof.

In the first case, inhibition might result in a clear NBR; in the second and third cases, it might reflect the local metabolic increases induced by the unaffected input and its ongoing processing, resulting in fMRI
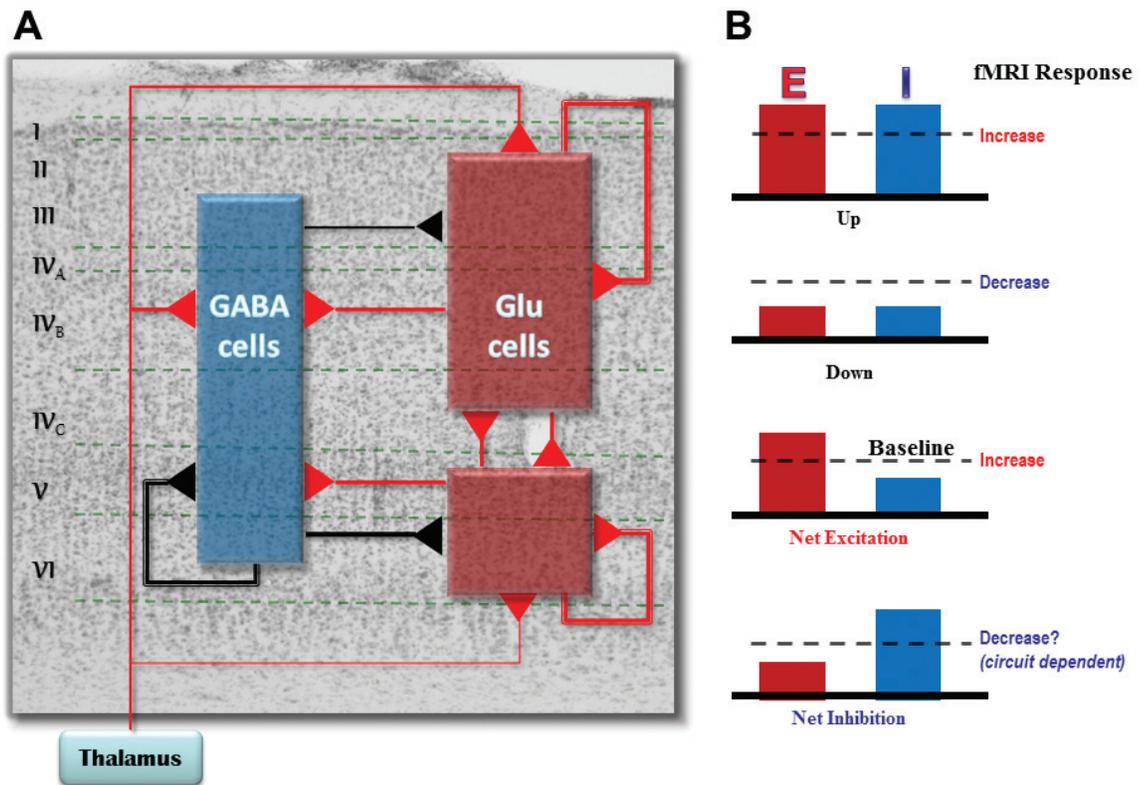


**Figure 2.** Principles of E-I circuits. ***A***, Model of a canonical cerebral microcircuit. Three neuronal populations interact with each other: supragranular-granular and infragranular pyramidal neurons, and GABA-ergic cells. Excitatory synapses are shown in red and inhibitory synapses in black. All groups receive excitatory thalamic input. Line width indicates the strength of connection. The circuit is characterized by the presence of weak thalamic input and strong recurrence. ***B***, Potential proportional and opposite-direction changes of cortical excitation and inhibition. Responses to large sustained input changes may occur while maintaining a well-balanced excitation (E) and inhibition (I) (up and down). The commonly assumed net excitation or inhibition might occur when the afferents drive the overall E-I balance in opposite directions. The balanced, proportional changes in E-I activity, which occur as a result of neuromodulatory input, are likely to strongly drive the hemodynamic responses. Glu, glutamatergic. Logothetis et al. (2008), their Fig. 2, and adapted from Douglas et al. (1989), reprinted with permission. Copyright ©2008 Macmillan Publishers.

activations. The fMRI responses might further blur the origin of inhibition, owing to the direct effects of the latter on the arterioles and microvessels.

## Neurophysiological Correlates of BOLD

At any given time, the active regions of a discharging neuron's membranes are considered to act as a current sink, whereas the inactive ones act as a current source for the active regions (Logothetis, 2008). The linear superposition of currents from all sinks and sources forms the extracellular field potential (EFP) measured by microelectrodes. The EFP captures at least three different types of EIN activity:

- Single unit activity (SUA), representing the action potentials of well-isolated neurons next to the electrode tip;

- Multiple unit activity (MUA), reflecting the spiking of small neural populations in a 100–300 μm radius sphere; and

- Perisynaptic activity of a neural population within 0.5–3.0 mm of the electrode tip, which is reflected in the variation of the low-frequency components of the EFP.

MUA and local field potential (LFP) can be reliably segregated by frequency band separation. A high-pass filter cutoff of 500 Hz is used in most recordings to obtain the MUA, and a low-pass filter cutoff of ~250 Hz is used to obtain LFP. A large number of experiments have presented data indicating that such a band separation does indeed underlie different neural events (Logothetis, 2008).

LFP signals and their different band-limited components (e.g., alpha, beta, gamma) are invaluable for understanding cortical processing because they are the only signs of integrative EIN processes. The relationship of neocortical LFPs and spiking activity to the BOLD signal itself has been examined directly in concurrent electrophysiology and fMRI experiments in the visual system of anesthetized (Logothetis et al., 2001) and alert (Goense and Logothetis, 2008) monkeys. These studies found that the BOLD responses reflect input and intracortical processing rather than pyramidal cell output activity. At first glance, both LFPs and spiking seemed to correlate with the BOLD response; subsequently, quantitative analysis indicated that LFPs are somewhat better predictors of the BOLD response than multiunit or single-unit spiking. The decisive finding leading to the papers' conclusion, however, was not the degree of correlation between the neural and the fMRI responses or the differential contribution that any type of signal made to the BOLD responses (Logothetis et al., 2001). Rather, it was the striking, undiminished hemodynamic responses in cases where spiking was entirely absent despite clear, strong stimulus-induced modulation of the field potentials (Logothetis and Wandell, 2004). Similar dissociations between spiking activity and the hemodynamic response had been demonstrated in earlier and very recent studies using other techniques (Mathiesen et al., 1998; Viswanathan and Freeman, 2007; Rauch et al., 2008).

## Conclusions

The limitations of fMRI are not related to physics or poor engineering and are unlikely to be resolved by increasing the sophistication and power of the scanners. The limitations have neural origins. That is, the fMRI signal cannot easily differentiate between function-specific processing and neuromodulation or between bottom-up and top-down signals, and it may potentially confuse excitation and inhibition. Further, the magnitude of the fMRI signal cannot be quantified to accurately reflect differences among brain regions or among tasks within the same region. The origin of the latter problem is not our current inability to accurately estimate $CMRO_2$ (cerebral metabolic rate of oxygen) from the BOLD signal but the fact that hemodynamic responses are sensitive to the size of the activated population of neurons. This population size may change as the sparsity of neural representations varies spatially and temporally. In cortical regions in which stimulus-related or task-related perceptual or cognitive capacities are sparsely represented (e.g., instantiated in the activity of a very small number of neurons), volume transmission (which likely underlies different states of motivation, attention, learning, and memory) may dominate hemodynamic responses and make it impossible to deduce the exact role of the area in the task at hand. Neuromodulation is also likely to affect the ultimate spatiotemporal resolution of the fMRI signal. Thus, the limitations of fMRI derive from the circuitry and functional organization of the brain as well as inappropriate experimental protocols that ignore this organization.

## Acknowledgment

NOTES

# References

Ackermann RF, Finch DM, Babb TL, Engel J, Jr (1984) Increased glucose metabolism during long-duration recurrent inhibition of hippocampal pyramidal cells. J Neurosci 4:251–264.

Douglas RJ, Martin KAC, Whitteridge D (1989) A canonical microcircuit for neocortex. Neural Comput 1:480–488.

Douglas RJ, Martin KA (2004) Neuronal circuits of the neocortex. Annu Rev Neurosci 27:419–451.

Goense JBM, Logothetis NK (2008) Neurophysiology of the BOLD fMRI signal in awake monkeys. Curr Biol 18:631–640.

Guillery RW, Sherman SM (2002) The thalamus as a monitor of motor outputs. Philos Trans R Soc Lond B Biol Sci 357:1809–1821.

Jueptner M, Weiller C (1995) Review: Does measurement of regional cerebral blood flow reflect synaptic activity? Implications for PET and fMRI. Neuroimage 2:148–156.

Jueptner M, Weiller C (1998) A review of differences between basal ganglia and cerebellar control of movements as revealed by functional imaging studies. Brain 121:1437–1449.

Logothetis NK (2002) The neural basis of the blood-oxygen-level-dependent functional magnetic resonance imaging signal. Phil Trans R Soc Lond B 357:1003–1037.

Logothetis NK (2003) The underpinnings of the BOLD functional magnetic resonance imaging signal. J Neurosci 23:3963–3971.

Logothetis NK (2008) What we can do and what we cannot do with fMRI. Nature 453:869–878.

Logothetis NK, Pauls JM, Augath MA, Trinath T, Oeltermann A (2001) Neurophysiological investigation of the basis of the fMRI signal. Nature 412:150–157.

Logothetis NK, Pfeuffer J (2004) On the nature of the BOLD fMRI contrast mechanism. Magn Reson Imaging 22:1517–1531.

Logothetis NK, Wandell BA (2004) Interpreting the BOLD signal. Ann Rev Physiol 66:735–769.

Luck SJ, Chelazzi L, Hillyard SA, Desimone R (1997) Neural mechanisms of spatial selective attention in areas V1, V2, and V4 of macaque visual cortex. J Neurophysiol 77:24–42.

Mathiesen C, Caesar K, Akgoren N, Lauritzen M (1998) Modification of activity-dependent increases of cerebral blood flow by excitatory synaptic activity and spikes in rat cerebellar cortex. J Physiol 512:555–566.

Motter BC (1993) Focal attention produces spatially selective processing in visual cortical areas V1, V2, and V4 in the presence of competing stimuli. J Neurophysiol 70:909–919.

Nudo RJ, Masterton RB (1986) Stimulation-induced [14C]2-deoxyglucose labeling of synaptic activity in the central auditory system. J Comp Neurol 245:553–565.

Rauch A, Rainer G, Logothetis NK (2008) The effect of a serotonin-induced dissociation between spiking and perisynaptic activity on BOLD functional MRI. Proc Natl Acad Sci USA 105:6759–6764.

Sherman SM (2005) Thalamic relays and cortical functioning. Cortical function: A view from the thalamus. Prog Brain Res 149:107–126.

Shmuel A, Augath MA, Oeltermann A, Logothetis NK (2006) Negative functional MRI response correlates with decreases in neuronal activity in monkey visual area V1. Nat Neurosci 9:569–577.

Viswanathan A, Freeman RD (2007) Neurometabolic coupling in cerebral cortex reflects synaptic more than spiking activity. Nat Neurosci 10:1308–1312.