Pathological Neural Activity Can Be Regulated by Deep Brain Stimulation and Antiseizure BMIs

BMIs have been developed to help people with disorders involving pathological neural activity in the brain, such as Parkinson disease and epilepsy. People with Parkinson disease benefit by having hand and arm tremor reduced. At present, there is no cure for Parkinson disease, and many people become resistant to pharmacological treatments. A deep brain stimulator (DBS) can help these people by delivering electrical pulses to targeted areas in the brain to disrupt the aberrant neural activity.

DBS is controlled by a neurostimulator implanted in the chest, with wires to stimulating electrodes in deep brain nuclei (eg, the subthalamic nucleus). The nuclei are continuously stimulated with these electrodes in order to alter the aberrant neural activity. This method can often greatly reduce Parkinson disease–related tremor for years. A DBS applied to different brain areas can also help people with essential tremor, dystonia, chronic pain, major depression, and obsessive-compulsive disorder.

Millions of people experiencing epileptic seizures are currently treated with antiseizure medications or neurosurgery, both of which often result in incomplete or impermanent seizure reduction. Antiseizure BMIs have shown considerable promise for further improving quality of life. These fully implanted BMIs operate by continuously monitoring neural activity in a brain region determined to be involved with seizures. They identify unusual activity that is predictive of seizure onset and then respond within milliseconds to disrupt this activity by electrically stimulating the same or a different brain region. This closed-loop response can be fast enough that seizure symptoms are not felt and seizures do not occur.

Replacement Part BMIs Can Restore Lost Brain Processing Capabilities

BMIs are capable of restoring more than lost sensory or motor capabilities. They are, in principle, capable of restoring internal brain processing. Of the four categories of BMIs, this is the most futuristic. An example is a "replacement part" BMI. The central idea is that if enough is known about the function of a brain region, and if this region is damaged by disease or injury, then it may be possible to replace this brain region.

Once the normal input activity to a brain region is measured (see next section), the function of the lost brain region could then be modeled in electronic hardware and software, and the output from this substitute processing center would then be delivered to the next brain region as though no injury had occurred. This would involve, for example, reading out neural activity with electrodes, mimicking the brain region's computational functions with low-power microelectronic circuits, and then writing in electrical neural activity with stimulating electrodes.

This procedure might also be used to initiate and guide neural plasticity. A replacement part BMI that is currently being investigated focuses on restoring memory by replacing parts of the hippocampus that are damaged due to injury or disease. Another potential application would be to restore the lost functionality of a brain region damaged by stroke.

These systems represent the natural evolution of the BMI concept, a so-called "platform technology" because a large number of systems can be envisioned by mixing and matching various write-in, computational, and read-out components. The number of neurological diseases and injuries that BMIs should be able to help address ought to increase as our understanding of the functions of the nervous system and the sophistication of the technology continue to grow.

Measuring and Modulating Neural Activity Rely on Advanced Neurotechnology

Measuring and modulating neural activity involves four broad areas of electronic technologies applied to the nervous system (so-called neurotechnology). The first area is the type of neural sensor; artificial neural sensors are designed with different levels of invasiveness and spatial resolution (Figure 39–2). Sensors that are external to the body, such as an *electroencephalogram* (EEG) cap, have been used extensively in recent decades. The EEG measures signals from many small metal disks (electrodes) applied to the surface of the scalp across the head. Each electrode detects average activity from a large number of neurons beneath it.

More recently, implantable electrode-array techniques, such as subdural *electrocorticography* (ECoG) and finely spaced micro-ECoG electrodes, have been used. Since ECoG electrodes are on the surface of the brain and are thus much closer to neurons than EEG electrodes, ECoG has higher spatial and temporal resolution and thus provides more information with which to control BMIs.

Most recently, arrays of *penetrating intracortical electrodes*, which we focus on in this chapter, have been used. The intracortical electrode arrays are made of silicon or other materials and coated with biocompatible materials. The arrays are implanted on the surface of the brain, with the electrode tips penetrating 1 to 2 mm

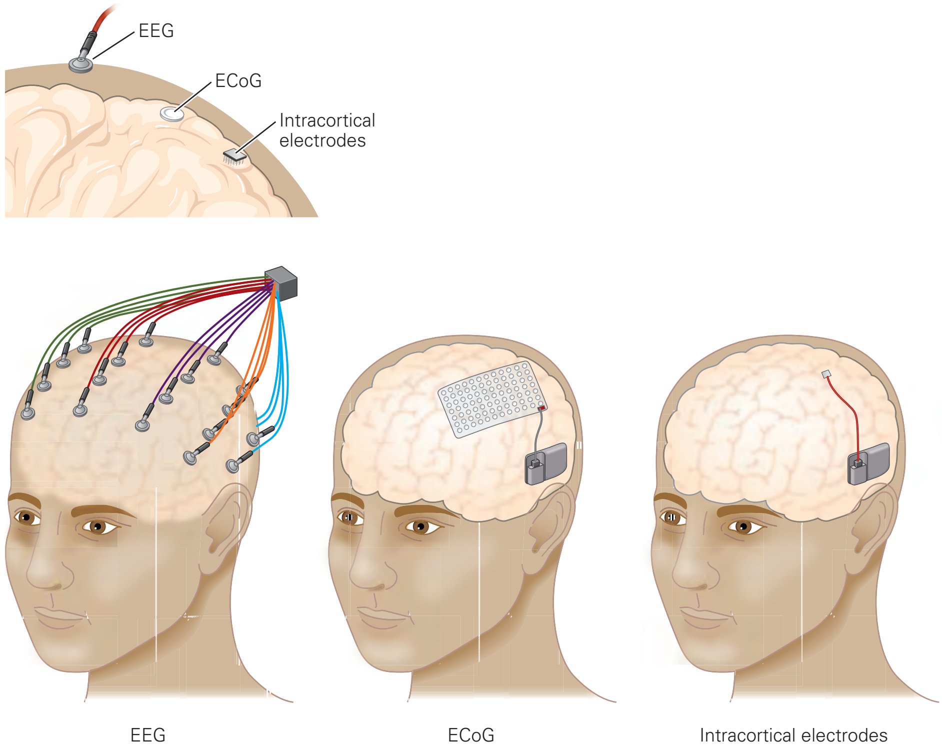


Figure 39–2 Brain–machine interfaces use different types of neural sensors. Electrical neural signals can be measured with various techniques ranging from electroencephalography (**EEG**) electrodes on the surface of the skin, to electrocorticography (**ECoG**) electrodes on the surface of the brain, to intracortical electrodes implanted in the outer 1 to 2 mm of cortex. The signals that can be measured range from the average of many neurons, to averages across fewer neurons, and finally to action potentials from individual neurons. (Adapted, with permission, from Blabe et al. 2015.)

into the cortex. They have the ability to record action potentials from individual neurons, as well as local field potentials from small clusters of neurons near each electrode tip. The electrodes are able to record high-fidelity signals because they are inserted into the brain, bringing the electrode tips within micrometers of neurons. This is beneficial for BMI performance because individual neurons are the fundamental information-encoding units in the nervous system, and action potentials are the fundamental units of the digital code that carries information from the input to the output region of a neuron. Moreover, intracortical electrodes can deliver electrical microstimulation to either disrupt neural activity (eg, DBS) or write in surrogate information (eg, proprioceptive or somatosensory information).

The second area of neurotechnology is scaling up the number of neurons measured at the same time. While one neuron contains some information about a person's intended movement, tens to hundreds of neurons are needed to move a BMI more naturally, and even more neurons are needed to approach naturalistic levels of motor function. Although it is possible to place electrode arrays in many areas across the brain, thereby gaining more information from multiple areas, a key challenge is to measure activity from thousands of neurons within each individual brain area. Many efforts are underway to achieve this goal, including use of electrode arrays with many tiny shafts, each with hundreds of electrode contacts along its length; many tiny electrodes that are not physically wired together, but are instead inserted into the brain as stand-alone islands that transmit data outside of the head and receive power wirelessly; and optical imaging technologies that can capture the activity of hundreds or more neurons by detecting how each neuron's fluorescence changes over time.

The third area is low-power electronics for signal acquisition, wireless data communications, and wireless powering. In contrast to the BMI systems described above, which implant a passive electrode array in which each electrode is wired to the outside world by a connector passing through the skin, future BMIs will be fully implanted like DBS systems. Electronic circuits are needed to amplify neural signals, digitize them, process them (eg, to detect when an action potential occurred or to estimate local field potential power), and transmit this information to a nearby receiver incorporated into a prosthetic arm, for example. Power consumption must be minimized for two reasons. First, the more power is consumed, the more power a battery or a wireless charging system would need to provide. Batteries would therefore need to be larger and replaced more often, and delivering power wirelessly is challenging. Second, using power generates heat, and the brain can only tolerate a small temperature increase before there are deleterious effects. These trade-offs are similar to those of smart phones, which represent the current best technology available for low-power electronics.

The final area is so-called supervisory systems. Software running on electronic hardware is at the heart of BMIs. Some software implements the mathematical operations of the neural decoding, while other software must tend to aspects of the BMI's overall operation. For example, the supervisory software should monitor whether or not a person wishes to use the prosthesis (eg, if the person is sleeping); if neural signals have changed, thereby requiring recalibration of the decoder; and overall BMI performance and safety.

Having discussed the range of different BMIs and neurotechnologies being developed, in the rest of this chapter we focus on motor and communication BMIs. We first describe different types of decoding algorithms and how they work. We then describe recent progress in BMI development toward assisting paralyzed people and amputees. Next, we consider how sensory feedback can improve BMI performance and how BMIs can be used as an experimental paradigm to address basic scientific questions about brain function. Finally, we conclude with a cautionary note about ethical issues that can arise with BMIs.

BMIs Leverage the Activity of Many Neurons to Decode Movements

Various aspects of movement—including position, velocity, acceleration, and force—are encoded in the activity of neurons throughout the motor system (Chapter 34). Even though our understanding of movement encoding in the motor system is incomplete, there is usually a reliable relationship between aspects of movement and neural activity. This reliable relationship allows us to estimate the desired movement from neural activity, a key component of a BMI.

To study movement encoding, one typically considers the activity of an individual neuron across repeated movements (referred to as "trials") to the same target. The activity of the neuron can be averaged across many trials to create a spike histogram for each target (Figure 39–3A). By comparing the spike histograms for different targets, one can characterize how the neuron's activity varies with the movement produced. One can also assess using the spike histograms whether the neuron is more involved in movement preparation or movement execution.

In contrast, estimating a subject's desired movement from neural activity (referred to as movement *decoding*) needs to be performed on an individual trial while the neural activity is being recorded. The activity of a single neuron cannot unambiguously provide such information. Thus, the BMI must monitor the activity of many neurons on a single trial (Figure 39–3B) rather than one neuron on many trials. A desired movement can be decoded from the neural activity associated with either preparation or execution of the movement. Whereas preparation activity is related to the movement goal execution activity is related to the moment-by-moment details of movement (Chapter 34).

Millions of neurons across multiple brain areas work together to produce a movement as simple as reaching for a cup. Yet in many BMIs, desired movements can be decoded reasonably accurately from the activity of dozens of neurons recorded from a single brain area. Although this may seem surprising, the fact is that the motor system has a great deal of redundancy—many neurons carry similar information about a desired movement (Chapter 34). This is reasonable because millions of neurons are involved in controlling the contractions of dozens of muscles. Thus, most of the neurons in regions of dorsal premotor cortex and primary motor cortex controlling arm movement are informative about most arm movements.

When decoding a movement, the activity of one neuron provides only incomplete information about the movement, whereas the activity of many neurons can provide substantially more accurate information about the movement. This is true for activity associated with both movement preparation and execution. There are two reasons why using multiple neurons is helpful for decoding. First, a typical neuron alone cannot unambiguously determine the intended movement direction. Consider a neuron whose activity (during either preparation or execution) is related to movement direction via a cosine function, known as a *tuning curve* (Figure 39–4A). If this neuron fires at 30 spikes per second, the intended movement direction could be either 120° or 240°. However, by recording from a second neuron whose tuning curve is different from that of the first neuron, the movement direction can be

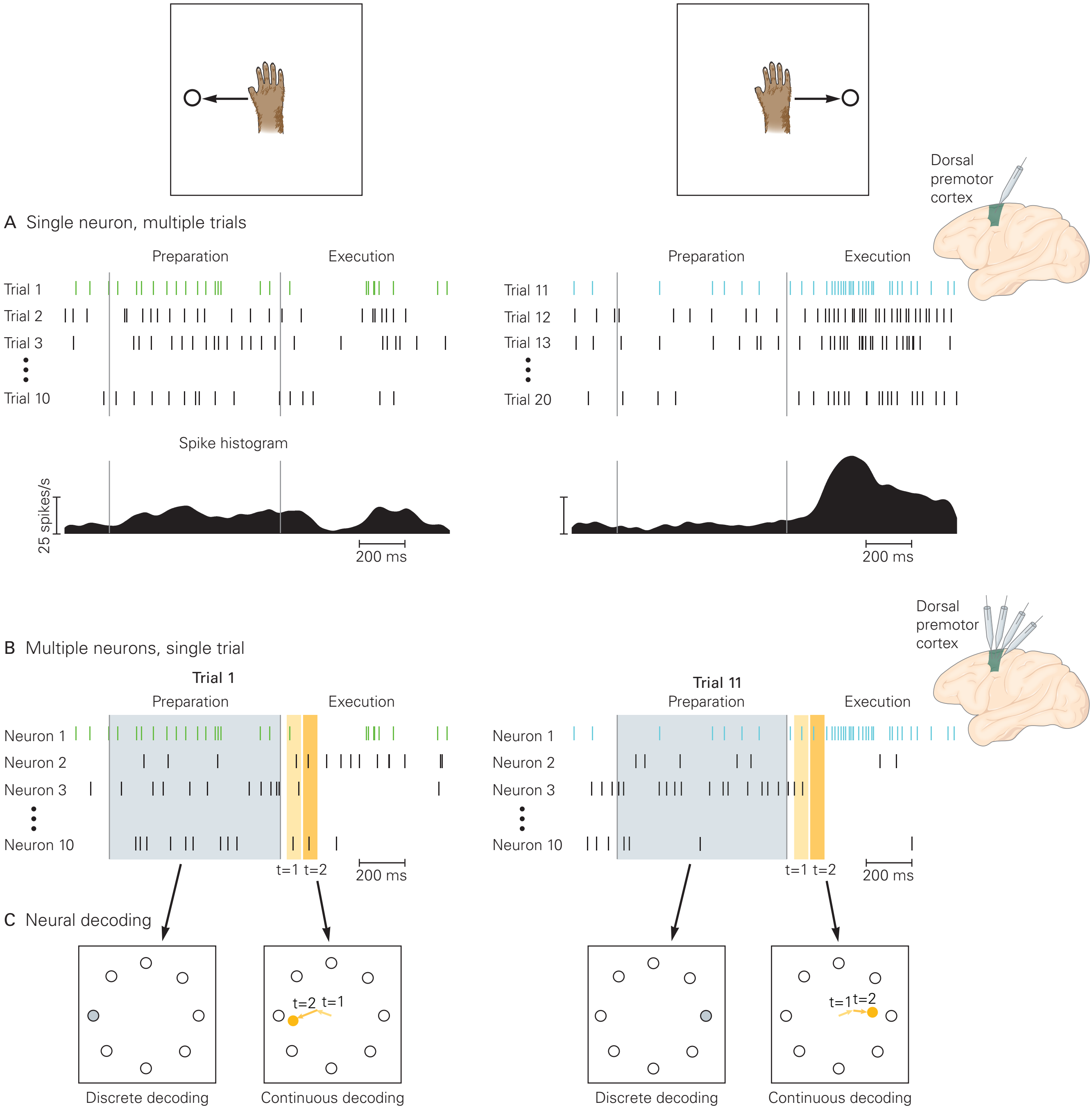


Figure 39–3 Movement encoding uses the activity of individual neurons averaged across experimental trials, whereas movement decoding uses the activity of many neurons on individual experimental trials.

A. Activity of one neuron recorded in the dorsal premotor cortex of a monkey preparing and executing leftward arm movements (*left*) and rightward arm movements (*right*). Characterizing the movement encoding of a neuron involves determining how the activity of the neuron on repeated leftward or rightward movements (each row of spike trains) relates to aspects of arm movement. **Below** is the spike histogram for this neuron for leftward and rightward movements, obtained by averaging neural activity across trials. This neuron shows a greater level of preparation activity for leftward movements and a greater level of execution activity for rightward movements. Many neurons in the dorsal premotor cortex and primary motor cortex show movement-related activity in both the preparation and execution epochs like the neuron shown.

B. Neural activity for many neurons recorded in the dorsal premotor cortex for one leftward movement (*left*) and one rightward movement (*right*). The spike trains for neuron 1 correspond to those shown in part **A**. Spike counts are taken during the preparation epoch, typically in a large time bin of 100 ms or longer to estimate movement goal. In contrast, spike counts are taken during the execution epoch typically in many smaller time bins, each lasting tens of milliseconds. Using such short time bins provides the temporal resolution needed to estimate the moment-by-moment details of the movement.

C. Neural decoding involves extracting movement information from many neurons on a single experimental trial. In the subject's workspace, there are eight possible targets (**circles**). Discrete decoding (see Figure 39–5) extracts the target location; the estimated target is filled in with **gray**. In contrast, continuous decoding (see Figure 39–6) extracts the moment-by-moment details of the movement; the **orange dot** represents the estimated position at one moment in time.

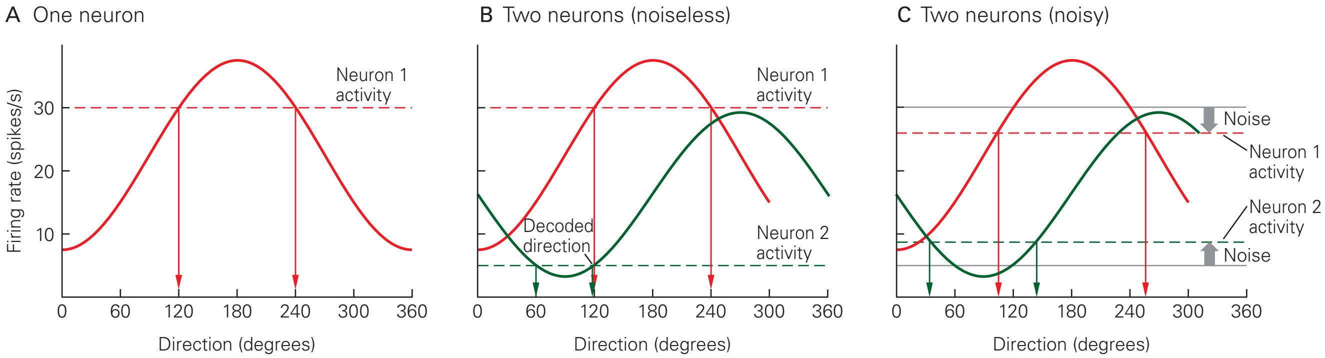


Figure 39–4 More than one neuron is needed for accurate movement decoding.

A. The tuning curve of one neuron defines how the neuron's activity varies with movement direction. If this neuron shows activity of 30 spikes/s, it could correspond to movement in the 120° or 240° direction.

B. A second neuron (**green**) with a different tuning curve shows activity of 5 spikes/s, which could correspond to movement in the 60° or 120° direction. The only movement direction consistent with the activity of both neurons is 120°, which is determined to be the decoded direction.

C. Because neural activity is "noisy" (represented as a vertical displacement of the dashed lines), it is usually not possible to conclusively determine the movement direction from the activity of two neurons. Here, no one movement direction is consistent with the activity of both neurons.

more accurately determined. If the second neuron fires at 5 spikes per second, corresponding to a movement in either the 60° or 120° direction, the only movement direction that is consistent between the two neurons is 120° (Figure 39–4B). Thus, by recording from these two neurons simultaneously, the intended reach direction can be determined more accurately than by recording from one neuron. (However, two neurons do not necessarily provide a perfect estimate of the intended reach direction due to noise, as described next.)

The second reason why decoding a movement from the activity of several neurons gives greater accuracy is because a neuron's activity level usually varies across repeated movements in the same direction. This variability is typically referred to as spiking "noise." Let us say that due to spiking noise the first neuron fires at slightly less than 30 spikes per second and the second neuron fires at slightly more than 5 spikes per second (Figure 39–4C). Under these conditions, no single movement direction is consistent with the activity level of both neurons. Instead, a compromise must be made between the two neurons to determine a movement direction that is as consistent as possible with their activities. By extending this concept to more than two neurons, the movement direction can be decoded even more accurately as the number of neurons increases.

Decoding Algorithms Estimate Intended Movements From Neural Activity

Movement decoders are a central component of BMIs. There are two types of BMI decoders: discrete and continuous (Figure 39–3C). A *discrete decoder* estimates one of several possible movement goals. Each of these movement goals could correspond to a letter on a keyboard. A discrete decoder solves a classification problem in statistics and can be applied to either preparation activity or execution activity. A *continuous decoder* estimates the moment-by-moment details of a movement trajectory. This is important, for example, for reaching around obstacles or turning a steering wheel. A continuous decoder solves a regression problem in statistics and is usually applied to execution activity rather than preparation activity because the moment-by-moment details of a movement can be more accurately estimated from execution activity (Chapter 34).

Motor BMIs must produce movement trajectories as accurately as possible to achieve the desired movement and typically use a continuous decoder to do this. In contrast, communication BMIs are concerned with enabling the individual to transmit information as rapidly as possible. Thus, the speed and accuracy with which movement goals (or keys on a keyboard) can be selected are of primary importance. Communication BMIs can use a discrete decoder to directly select a desired key on a keyboard or a continuous decoder to continuously guide the cursor to the desired key, where only the key eventually struck actually contributes to information conveyance. This seemingly subtle distinction has implications that influence the type of neural activity required and therefore the brain area that is targeted, as well as the type of decoder that is used.

Neural decoding involves two phases: calibration and ongoing use. In the calibration phase, the relationship between neural activity and movement is characterized by a statistical model. This can be achieved by recording neural activity while a paralyzed person attempts to move, imagines moving, or passively observes movements of a computer cursor or robotic limb. Once the relationship has been defined, the statistical model can then be used to decode new observed neural activity (ongoing use phase). The goal during the ongoing use phase is to find the movement that is most consistent with the observed neural activity (Figure 39–4B,C).

Discrete Decoders Estimate Movement Goals

We first define a population activity space, where each axis represents the firing rate of one neuron. On each trial (ie, movement repetition), we can measure the firing rate of each neuron during a specified period, and together they yield one point in the population activity space. Across many trials, involving multiple movement goals, there will be a scatter of points in the population activity space. If the neural activity is related to the movement goal, then the points will be separated in the population activity space according to the movement goal (Figure 39–5A). During the calibration phase, *decision boundaries* that partition the population activity space into different regions are determined by a statistical model. Each region corresponds to one movement goal.

During the ongoing use phase, we measure new neural activity for which the movement goal is unknown (Figure 39–5B). The decoded movement goal is determined by the region in which the neural activity lies. For example, if the neural activity lies within the region corresponding to the leftward target, then the discrete decoder would guess that the subject intended to move to the leftward target on that trial. It is possible that the subject intended to move to the rightward target, even though the recorded activity lies within the region corresponding to the leftward target. In this case, the discrete decoder would incorrectly estimate the subject's intended movement goal. Decoding accuracy typically increases with an increasing number of simultaneously recorded neurons.

Continuous Decoders Estimate Moment-by-Moment Details of Movements

Arm position, velocity, acceleration, force, and other aspects of arm movement can be decoded using the methods described here with varying levels of accuracy. For concreteness, we will discuss decoding movement velocity because it is one of the quantities most strongly reflected in the activity of motor cortical

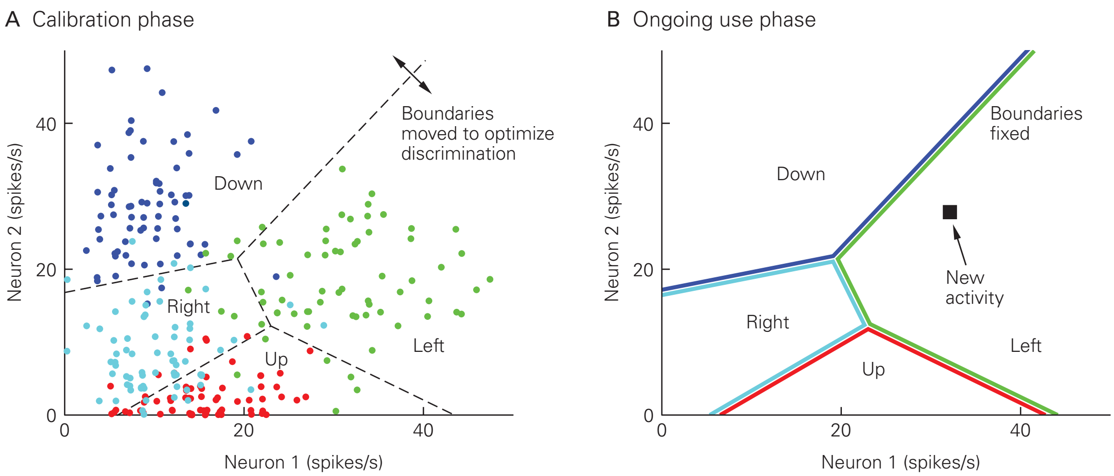


Figure 39–5 Discrete decoding.

A. Calibration phase. A population activity space is shown for two neurons, where each axis represents the firing rate of one neuron. On each trial (ie, movement repetition), the activity of the two neurons together defines one point in the population activity space. Each point is colored by the movement goal, which is known during the calibration phase. Decision boundaries (**dashed lines**) are determined by a statistical model to optimize discrimination among the movement goals. The decision boundaries define a region in the population activity space for each movement goal.

B. Ongoing use phase. During this phase, the decision boundaries are fixed. If we record new neural activity (**square**) for which the movement goal is unknown, the movement goal is determined by the region in which the neural activity lies. In this case, the neural activity lies in the region corresponding to the leftward target, so the decoder would guess that the subject intended to move to the leftward target.

neurons and is the starting point for the design of most BMI systems.

Consider a population of neurons whose level of activity indicates the movement velocity (ie, speed and direction). During the calibration phase, a "pushing vector" is determined for each neuron (Figure 39–6A). A pushing vector indicates how a neuron's activity influences movement velocity. Various continuous decoding algorithms differ in how they determine the pushing vectors. One of the earliest decoding algorithms, the population vector algorithm (PVA), assigns each neuron's pushing vector to point along the neuron's preferred direction (see Figure 34–22A). A neuron's preferred direction is defined as the direction of movement for which the neuron shows the highest level of activity (ie, peak of curves in Figure 39–4). Much of the pioneering work on BMIs used the PVA. However, the PVA does not take into account the properties of the spiking noise (ie, its variance and covariance across neurons), which influences the accuracy of the decoded movements. A more accurate decoder, the optimal linear estimator (OLE), incorporates the properties of the spiking noise to determine the pushing vectors.

During the ongoing use phase, the pushing vectors are each scaled by the number of spikes emitted by the corresponding neuron at each time step (Figure 39–6B). At each time step, the decoded movement is the vector sum of the scaled pushing vectors across all neurons. The decoded movement represents a change in position during one time step (ie, velocity). The BMI cursor (or limb) position (Figure 39–6C) is then updated according to the decoded movement.

To further improve decoding accuracy, the estimation of velocity at each time step should take into account not only current neural activity (as illustrated in Figure 39–6), but also neural activity in the recent past. The rationale is that movement velocity (and other kinematic variables) changes gradually over time, and so neural activity in the recent past should be informative about the movement velocity. This can be achieved by temporally smoothing the neural activity before applying a PVA or OLE or by using a Kalman filter to define a statistical model describing how movement velocity (or other kinematic variables) changes smoothly over time. With a Kalman filter, the estimated velocity is a combination of the scaled pushing vectors at the current time step (as in Figure 39–6B) and the estimated velocity at the previous time step. Indeed, continuous decoding algorithms that take into account neural activity in the recent past have been shown to provide higher decoding accuracy than those that do not. The Kalman filter and its extensions are widely used in BMIs and among the most accurate continuous decoding algorithms available.

Increases in Performance and Capabilities of Motor and Communication BMIs Enable Clinical Translation

Patients with paralysis wish to perform activities of daily living. For people with ALS or upper spinal cord injury who are unable to speak or to move their arms, the most desired tasks are often the ability to communicate, to move a prosthetic (robotic) arm, or to move the paralyzed arm by stimulating the musculature. Having described how neural signals can be read out from motor areas of the brain and how these electrical signals can be decoded to arrive at BMI control signals, we now describe recent progress toward restoring these abilities.

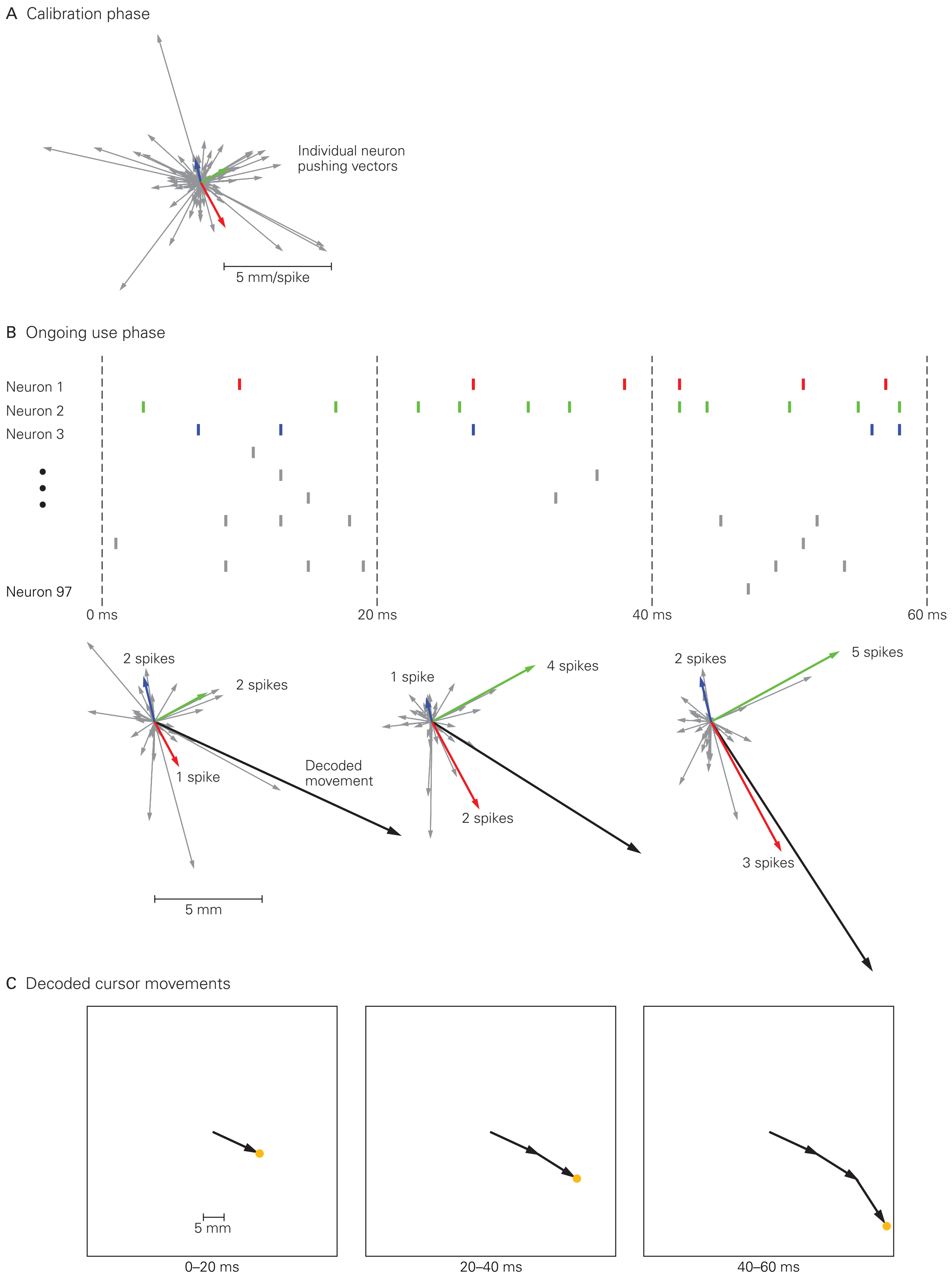
The majority of laboratory studies are carried out in able-bodied nonhuman primates, although paralysis is sometimes transiently induced in important control experiments. Three types of experimental paradigms are in broad use, differing in the exact way in which arm behavior is instructed and visual feedback is provided during BMI calibration and ongoing use. Setting

Figure 39–6 (Opposite) Continuous decoding.

A. During the calibration phase, a pushing vector is determined for each of 97 neurons. Each vector represents one neuron and indicates how one spike from that neuron drives a change in position per time step (ie, velocity). Thus, the units of the plot are millimeters per spike during one time step. Different neurons can have pushing vectors of different magnitudes and directions.

B. During ongoing use, spikes are recorded from the same neurons as in panel **A** during movement execution. At each time step, the new length of an arrow is obtained by starting with its previous length in panel **A** and scaling it by the number of spikes produced by the neuron of the same color during that time step. If a neuron does not fire, there is no arrow for that neuron during that time step. The decoded movement (**black arrow**) is the vector sum of the scaled pushing vectors, representing a change in position during one time step (ie, velocity). For a given neuron, the direction of its scaled pushing vectors is the same across all time steps. However, the magnitudes of the scaled pushing vectors can change from one time step to the next depending on the level of activity of that neuron.

C. The decoded movements from panel **B** are used to update the position of a computer cursor (**orange dot**), robotic limb, or paralyzed limb at each time step.


A Calibration phase
Individual neuron pushing vectors
5 mm/spike
B Ongoing use phase
Neuron 1
Neuron 2
Neuron 3
Neuron 97
0 ms
20 ms
40 ms
60 ms
2 spikes
2 spikes
1 spike
Decoded movement
5 mm
1 spike
4 spikes
2 spikes
2 spikes
5 spikes
3 spikes
C Decoded cursor movements
5 mm
0–20 ms
20–40 ms
40–60 ms

these differences aside, we focus below on how BMIs function and perform. We also highlight recent pilot clinical trials with people with paralysis.

Subjects Can Type Messages Using Communication BMIs

To investigate how quickly and accurately a communication BMI employing a discrete decoder and preparation activity can operate, monkeys were trained to fixate and touch central targets and prepare to reach to a peripheral target that could appear at one of several different locations on a computer screen. Spikes were recorded using electrodes implanted in the premotor cortex. The number of spikes occurring during a particular time window during the preparation epoch was used to predict where the monkey was preparing to reach (Figure 39–7A). If the decoded target matched the peripheral target, a liquid reward was provided to indicate a successful trial.

By varying the duration of the period in which spike counts are taken and the number of possible targets, it was possible to assess the speed and accuracy of target selections (Figure 39–7B). Decoding accuracy tended to increase with the period in which spike counts are taken because spiking noise is more easily averaged out in longer periods.

An important metric for efficient communication is information transfer rate (ITR), which measures how much information can be conveyed per unit time. A basic unit of information is a bit, which is specified by a binary value (0 or 1). For example, with three bits of information, one can specify which of $2^3 = 8$ possible targets or keys to press. Thus, the metric for ITR is bits per second (bps). ITR increases with the period in which spike counts are taken, then declines. The reason is that ITR takes into account both how accurately and how quickly each target is selected. Beyond some point of diminishing returns of a longer period, accuracy fails to increase rapidly enough to overcome the slowdown in target-selection rate accompanying a longer period.

Overall performance (ITR) increases with the number of possible targets, despite a decrease in decoding accuracy, because each correct target selection conveys more information. Fast and accurate communication

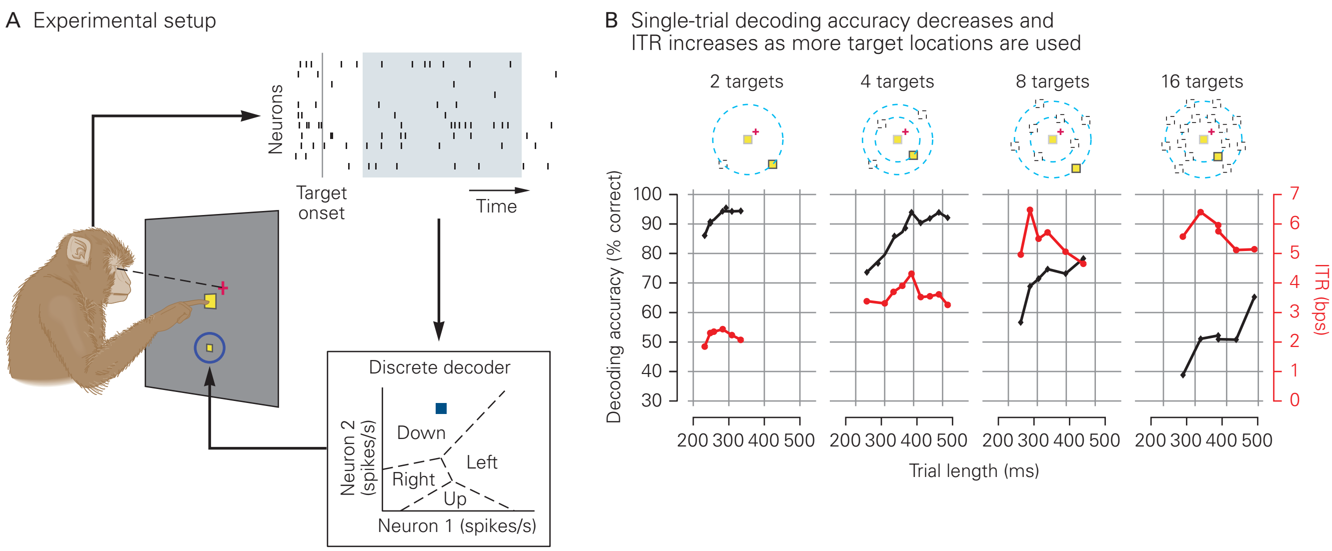


Figure 39–7 A communication brain–machine interface can control a computer cursor using a discrete decoder based on neural activity during the preparation epoch.

A. After a monkey touched a central target (**large yellow square**) and fixated a central point (**red +**), a peripheral target (**small yellow square**) appeared and the monkey prepared to reach to it. Spike counts were taken during the preparation epoch and fed into a discrete decoder. The duration of the period in which spike counts are taken (ie, width of **light blue** shading) affects decoding performance and information transfer rate (**ITR**) (see panel **B**). Based on the spike counts (**blue square**), the discrete decoder guessed the target the monkey was preparing to reach to.

B. Decoding accuracy (**black**) and information transfer rate (**ITR, bits/s; red**) are shown for different trial lengths and numbers of targets. Trial length was equal to the duration of the period in which spike counts were taken (varied during the experiment) plus 190 ms (fixed during the experiment). The latter provided time for visual information of the peripheral target to reach the premotor cortex (150 ms), plus the time to decode the target location from neural activity and render the decoded target location on the screen (40 ms). (Adapted, with permission, from Santhanam et al. 2006.)