Review

Central nervous system physiology

John Rothwell a,*, Andrea Antal b, David Burke c, Antony Carlsen d, Dejan Georgiev e, Marjan Jahanshahi f, Dagmar Sternad g, Josep Valls-Solé h, Ulf Ziemann i

a Department of Clinical and Movement Neuroscience, UCL Queen Square Institute of Neurology, London, UK
b Department of Neurology, University Medical Center Göttingen, Germany
c Department of Neurology, Royal Prince Alfred Hospital, University of Sydney, Sydney 2050, Australia
d School of Human Kinetics, University of Ottawa, Ottawa, Canada
e Department of Neurology, University Medical Centre Ljubljana, Slovenia
f Department of Clinical and Movement Neuroscience, UCL Queen Square Institute of Neurology, London, UK
g Departments of Biology, Electrical & Computer Engineering, and Physics, Northeastern University, Boston, MA 02115, USA
h Institut d’Investigació Biomèdica August Pi i Sunyer, Villarroel, 170, Barcelona, Spain
i Department of Neurology and Stroke, and Hertie Institute for Clinical Brain Research, Eberhard Karls University, Tübingen, Germany

This review is the second in the series of articles on the use of clinical neurophysiology for the study of movement disorders. It focuses on the most useful non-invasive methods and techniques that can probe the neurophysiology of the central nervous system in humans. Tools include reflex studies, transcranial brain stimulation, electroencephalography and reaction times.

This is the second chapter of the series on the use of clinical neurophysiology for the study of movement disorders. It focusses on methods that can be used to probe neural circuits in brain and spinal cord. These include use of spinal and supraspinal reflexes to probe the integrity of transmission in specific pathways; transcranial methods of brain stimulation such as transcranial magnetic stimulation and transcranial direct current stimulation, which activate or modulate (respectively) the activity of populations of central neurones; EEG methods, both in conjunction with brain stimulation or with behavioural measures that record the activity of populations of central neurones; and pure behavioural measures that allow us to build conceptual models of motor control. The methods are discussed mainly in relation to work on healthy individuals. Later chapters will focus specifically on changes caused by pathology.

© 2021 International Federation of Clinical Neurophysiology. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
1. Introduction

The basic techniques of clinical neurophysiology were summarised in the previous chapter. Here we review how to apply these methods to study circuits in the brain and spinal cord. The methods are described with reference to studies on healthy individuals; the application to specific types of movement disorder will be described in subsequent chapters. The first sections describe how reflex studies can give information on the excitability and transmission in spinal and supraspinal reflex pathways. These pathways, and the information that can be obtained from them are relatively well-defined. They contrast with the less specific methods of transcranial brain stimulation and EEG, which give information about the excitability and activity of mixed populations of neurones. The final level of investigation is at the behavioural level, where data can inform high-level conceptual models of how the brain might control movement.

2. Spinal cord reflexes including long latency reflexes

The state of spinal cord circuitry is critical for any limb movement, whether voluntary or involuntary: the brain may command a specific movement, but what actually occurs is only that which the spinal cord allows. The motor command can be modified reflexly by sensory feedback at multiple levels of the neuraxis. Equally however the circuitry of the spinal cord is subject to supraspinal controls, some conscious and voluntary, some automatic and involuntary. All reflex pathways in the spinal cord are subject to these descending influences, exerted on interneurons associated with that pathway. Even the monosynaptic group Ia pathway can be controlled from above, by modulating the activity of the presynaptic inhibitory interneuron. A limited number of pathways can be investigated reliably in human subjects (Fig. 1; for techniques see Pierrot-Deseilligny and Burke, 2012). This discussion will be confined to some of the mecha-
nisms operating at spinal level and some operating at cortical level.

It is prudent to remember that, in humans, the conduction velocities of all classes of afferent fibres cover a wide range and that those of the fastest cutaneous afferents overlap virtually completely those of the fastest muscle afferents (Macefield et al., 1989). In addition, the reflex effects, particularly the responses to mechanical stimuli, are produced through a number of interacting pathways, at a number of levels of the neuraxis. While it has been convenient to use electrical stimulation and abrupt perturbations to separate out the different components of the overall response, the resulting circuit diagram tells one little about how those circuits normally function. Similarly, it may be convenient to study reflex pathways in subjects at rest, but this belies the fact that the motor system exists for movement.

2.1. Spinal reflex mechanisms

In studies using transcranial stimulation of the motor cortex, a cortical locus of action can be inferred if there is no significant alteration in spinal reflex function. Traditionally, this has involved testing the excitability of the target motoneuron pool using H reflexes or F waves. While the H reflex constitutes a better measure of the excitability of the motoneuron pool, both approaches have flaws, as discussed below. An alternative has been to record the change in the descending corticospinal volleys produced by TMS (Di Lazzaro and Rothwell, 2014), but this is practical only in those patients with implanted epidural electrodes. In addition, it is uncertain whether the corticospinal volleys target the relevant motoneuron pool. When these techniques provide complementary data one can have a reasonable certainty about the site of action. Nevertheless, at the present time, there are no techniques for stimulating the motor cortex that do not also have direct or indirect effects on spinal cord circuitry.

2.1.1. The H reflex

The H reflex depends on monosynaptic excitation of α motoneurons by group Ia afferents from muscle spindles. Muscle spindle-like structures have been identified in all skeletal muscles, except the facial muscles and the digastric. It is therefore not surprising that H reflexes can be recorded for multiple muscles throughout the body, particularly when the circuitry is potentiated.

---

Fig. 1. Some spinal circuits that can be tested reliably in human subjects. Group Ia circuits in red; group II circuits in blue; descending controls in green. IN: interneuron. MN: motoneuron. PAD INs: interneurones that produce primary afferent depolarisation and are, thereby, responsible for presynaptic inhibition. HD: homosynaptic depression (post-activation depression of transmitter release at the synapse). NA: brainstem noradrenergic pathway (suppressing group II reflex excitation). RC: Renshaw cell. From (Pierrot-Deseilligny and Burke, 2012) with permission.
by a background voluntary contraction (Burke, 2016). The reflex response can then be identified as a change in the probability of discharge of voluntarily active single motor units, or as a peak at the appropriate latency in the averaged electromyography (EMG) of the contracting muscle (Fig. 2A, B). When a reflex pathway contains a number of synapses, conduction is dispersed and the response is usually polyphasic. This is best quantified by averaging full-wave rectified EMG rather than raw EMG. However, this does not apply to the H reflex because the connectivity of the excitation is largely monosynaptic, and the response biphasic. Because of this, H-reflexes are usually measured peak-to-peak.

The H reflex is often called a “monosynaptic reflex”, but this ignores the fact that whether motoneurons can discharge reflexly depends on the balance between monosynaptic excitation and disynaptic inhibition. The electrical stimulus for the H reflex will excite Ib afferents from Golgi tendon organs (and also cutaneous afferents from appropriate skin regions). Because of the dispersion of the la volley at the motoneuron pool, disynaptic inhibition due to the Ib volley can curtail the group la excitation even though it arrives 0.5–1.0 ms after the onset of la excitation. This has been demonstrated in the cat (Araki et al., 1960) and for the H reflex of human subjects (Burke et al., 1984; Marchand-Pauvert et al., 2002). Presynaptic inhibition due to an axo-axonal synapse on the presynaptic terminals of the la afferent (“PAD INs” in Fig. 1) represents an additional mechanism through which the group la excitatory post-synaptic potential (EPSP) can be attenuated, thus reducing the reflex discharge (see Chapter 8 in (Pierrot-Deseilligny and Burke, 2012)). Valid techniques have been developed to document presynaptic inhibition for different motoneuron pools (e.g., (Aymard et al., 2000; Berardelli et al., 1987; Faist et al., 1996; Hultborn et al., 1987; Meunier and Pierrot, 1989)).

In testing the excitability of the motoneuron pool using a peripheral (reflex) input, it is important that the recruitment of motoneurons in the test reflex mirrors the recruitment sequence of the corticospinal drives associated with a voluntary contraction or TMS. The available evidence supports this requirement for the H reflex: volition, TMS and group la excitation appear to access motoneurons in the same orderly sequence, related to the size of the motoneuron, the lowest-threshold motoneurons being the smallest in the pool.

2.1.1. Limitations of the H reflex. In subjects who are relaxed, H reflexes can be recorded reliably only from soleus, quadriceps femoris and flexor carpi radialis. They can usually be recorded at rest from muscles such as biceps brachii, but technical limitations commonly interfere with the recording. H reflexes can occur in the intrinsic muscles of the hand, even at rest (Trontelj, 1973), but this usually requires a strong stimulus and it is difficult to distinguish reflex responses from F waves. H reflexes are more clearly seen during a steady voluntary contraction (see Fig. 2A,B). A voluntary contraction raises the motoneuron pool to firing threshold and

Fig. 2. H reflexes of abductor pollicis brevis during a voluntary contraction and F waves of the thenar muscles in the same subject. A, H reflex recorded using unrectified electromyography (EMG) (4 averages, each of 32 sweeps), as used to calculate H-reflex latencies. B, H reflex recorded using rectified EMG (3 averages, each of 32 sweeps). C, F waves of abductor pollicis brevis (APB) (3 averages, each of 32 sweeps). Note the small M wave in A and B, and the maximal M wave in C. The F wave latency is slightly shorter than the H reflex latency (on average the latency of the H reflex of APB = \(-1.1 \times \) latency of fastest F wave. From (Espiritu et al., 2003), with permission.
Fig. 3. Inability of the H reflex to detect reflex actions mediated through interneurons. Panel (a) shows the circuitry proposed to explain the data in panel (b). Stimulation of the superficial (cutaneous) branch of the radial nerve at the wrist inhibits propriospinal neurons located at the C3-C4 level, thus reducing the component of the corticospinal volley transmitted to the motoneuron pool of extensor carpi radialis (ECR). Panel (b) shows that cutaneous afferents in the radial nerve suppressed the background electromyography (EMG) of a steady voluntary contraction of ECR (filled circles) and the motor evoked potential (MEP) of the contracting ECR (filled triangles), but did not significantly suppress the H reflex of contracting ECR (open circles). The inhibition was therefore not at the motoneuron, and must have occurred at an interneuron, involving “disfacilitation”, rather than direct inhibition. The central delay, i.e., the extra interval spent within spinal cord circuitry, was 4 ms, which is too short for a pathway outside the spinal cord. This is consistent with transmission through propriospinal neurons located a few segments above the motoneuron pool. From (Pierrot-Deseilligny and Burke, 2012) with permission.

2.1.2. The F wave

In clinical practice, F waves are elicited by supramaximal stimuli, thus avoiding contamination by the H reflex. Fig. 2D illustrates F waves for the thenar muscles of a healthy subject, with traces superimposed in the upper panel and as a raster in the lower panel. The F waves vary in latency, morphology and size because they derive from different motoneurons (though the most excitable motoneurons may appear more than once in a sequence, producing so-called “repeater F waves”). Because of slow decay of the large M wave, it is often convenient to record F waves using a high-pass filter of 30 or 100 Hz. This does not affect latencies significantly but reduces the amplitude for individual F waves. The higher filter setting is particularly useful when averaging rectified F waves to obtain a measure of F wave activity (as in Fig. 2C) because it is essential for the averaging that the trace returns to baseline between the M wave and the F responses.

A stimulus that is supramaximal for all motor axons will also generate an intense afferent discharge, which will excite low-threshold motoneurons in the pool through reflex pathways (Fig. 4). The H reflex produced in those motoneurons will not be seen because of occlusion by the antidromic volley in motor axons, but the collision between the reflex discharge and the antidromic...
voltage in motor axons will prevent the reflexly accessible motoneurons from generating F waves. Under these circumstances, the antidromic volley can access only those motoneurons that could not be discharged by the intense Ia afferent volley. This is an issue for the thenar muscles and, as a result, the conduction velocities for the slowest α motor axons innervating these muscles are not revealed by F-wave techniques.

2.1.2. Limitations of F waves. There are two important limitations on the use of F waves to measure the excitability of the motoneuron pool. First, F waves are relatively insensitive to change in motoneuron excitability (Burke, 2014; Hultborn and Nielsen, 1995). Secondly, F waves will probably be generated by the wrong motoneurons. They provide insight into only those motoneurons that could not be discharged reflexly (see above, Fig. 4). For the intrinsic muscles of the hand, these are mainly motoneurons of higher threshold, not those recruited preferentially by voluntary effort or TMS. F wave studies cannot provide an adequate control for the absence of a spinal locus of action in TMS studies, at least for upper limb muscles. This issue may be of lesser importance for tibialis anterior. Motoneurons of tibialis anterior have low excitability to group Ia inputs (even if they have relatively high excitability to corticospinal inputs), the opposite of the situation with soleus motoneurons.

2.2. Medium- and long-latency reflexes

Stretch reflex responses that have a longer latency than the spinal stretch reflex were first identified in the upper limb, and have been referred to as long-latency responses (LLRs). In the lower limb there may be more than one reflex component at long latency, and the terms medium-latency and long-latency have been used for them. These responses have different mechanisms (see below). Here the term "long-latency" will be used to refer to both.

Abrupt stretch of a muscle can generate three peaks of EMG activity. To demonstrate the second and third peaks, the raw EMG activity of a background contraction should be full-wave rectified before averaging. In the upper limb, the first peak is generally considered to be a spinal stretch reflex and the third peak to be automatic/voluntary, but there has been debate about the mechanisms underlying the second peak. The most popular theories have been (i) that this peak is generated by the same afferents as are responsible for the short-latency spinal response, the longer latency being due to a longer pathway within the neuraxis, to and from the cerebral cortex (e.g., (Cheney and Fetz, 1984; Marsden et al., 1973, 1976, 1977a, 1981a; Noth et al., 1985; Phillips, 1969)), and (ii) that the long-latency response is mediated by afferents with a slower conduction velocity, namely group II afferents from secondary spindle endings (Matthews, 1984). The second hypothesis has been largely discredited for the intrinsic muscles of the hand and for flexor pollicis longus (Matthews, 1991; Rothwell, 1990). The transcortical pathway to and from cortex involves the posterior columns and the corticospinal system (Marsden et al., 1977b, 1977c).

Single stimuli to the mixed median nerve at the wrist or to digital nerves decreases the excitability of motor cortex, a phenomenon referred to as short-latency afferent inhibition (Tokimura et al., 2000), but this is attenuated during voluntary contraction. The situation may be different with natural or repetitive afferent inputs. During a background contraction long-latency transcortical responses can be evoked by both cutaneous and muscle afferents in the intrinsic muscles of the hand (Deuschl et al., 1985). This is relevant for PAS (paired associative stimulation). Here plasticity of the motor cortex can be induced in subjects who are otherwise at rest by low-frequency repetitive stimulation.
(at 0.05 Hz) for ~30 min of both the mixed median nerve (or of the digital nerves) and of the motor cortex, but with the TMS delayed by an interval appropriate for conduction of the somatosensory volley to primary sensory cortex and thence to M1, ~25 ms (Stefan et al., 2000). Similarly, low-amplitude vibration of the hand can potentiate the MEP in muscles of the hand but not elsewhere (Rosenkranz and Rothwell, 2003). In accordance with these findings, the MEP can be modified by amputation and by cutaneous anaesthesia (Brasil-Neto et al., 1993; Cohen et al., 1991; Werhahn et al., 2001; Ziemann et al., 1998). These findings are consistent with a transcortical reflex with potent feedback from mechanoreceptors to cortex. That there is a transcortical pathway does not exclude the likelihood that the conditioning stimuli would have effects at spinal level as well.

Group II muscle afferents still play an important role in motor control, though this has not been fully appreciated because it is difficult to stimulate them mechanically or electrically without also activating group I afferents. Group II effects are suggested by latency, excessive delay with cooling, minimal effect of muscle vibration and suppression by tizanidine. There is considerable evidence that group II afferents contribute to the stretch reflex of the decerebrate cat (Matthews, 1969) and to reflex responses evoked by electrical stimulation of mixed nerves (Chapter 7 in (Pierrot-Deseilligny and Burke, 2012)). To date all investigated group II reflex effects have been facilitatory: there has been no evidence of a “flexor reflex afferent” pattern, with inhibition of extensor motoneurons, as seen in the acutely spinalised cat (Eccles and Lundberg, 1959). In the upper limb, it is likely that both long-loop Ia and spinal group II mechanisms contribute to long-latency EMG responses but in different proportions for different muscle groups (Lourenco et al., 2006; Thilmann et al., 1991).

In the lower limb the situation differs. Again, abrupt stretch produces three EMG peaks. The first corresponds to the spinal stretch reflex, the second is dependent on group II afferents (and convergent excitation from group I afferents on the group II interneuron (see Chapter 7 in (Pierrot-Deseilligny and Burke, 2012)), and the third is a transcortical reflex (Petersen et al., 1998). The latter may be responsible for the “functional stretch reflex” produced in leg muscles by perturbations of stance (Nashner, 1976). Many group II interneurons (also called “group I/II interneurons”) are located in segments caudal to motoneuron pools in low lumbar segments and are therefore “propriospinal”. The inputs and connections have been studied extensively using electrical stimuli (Pierrot-Deseilligny and Burke, 2012), and the functional importance of these circuits has been demonstrated using perturbations to stance in human subjects.

LLRs may be produced by electrical stimulation of either a mixed nerve or purely cutaneous afferents (e.g. (Deuschl et al., 1985; Lourenco et al., 2006; Noth et al., 1985)) or by natural stimuli such as abrupt stretch (e.g. (Marsden et al., 1973, 1976; Marsden et al., 1977b; Marsden et al., 1981b; Noth et al., 1985; Petersen et al., 1998)), or, with the lower limb, an abrupt perturbation to stance (Nashner, 1976; Schieppati and Nardone, 1997). Whether using natural or electric stimuli, the reflex responses usually require a background contraction, and should be quantified by measuring the area of the rectified EMG response above the background contraction level (or by measuring the change in discharge probability of single motor units (Lourenco et al., 2006)).

2.2.1. Limitations of LLRs

As mentioned above, group IA afferents have a low threshold and a faster conduction velocity than group II afferents, and it is therefore impossible to stimulate group II afferents selectively, whether doing so with electrical stimulation of a mixed nerve or muscle stretch. As a result, group II reflex patterns will appear in a motoneuron pool that has been conditioned by the group I spinal input. Similarly any long-loop component of the overall reflex response will occur in motoneurons that have been conditioned by spinal activity.

It is debatable whether studying medium- and long-latency responses are of diagnostic value in patients, given greater sensitivity and specificity of other diagnostic procedures, particularly radiology (multiple sclerosis: central pathology demonstrated by delayed or absent LLR in theer muscles relative to the short-latency spinal response (Deuschl et al., 1988)) and genetics (Huntington’s chorea: preserved H reflex but depressed LLR in hand muscles (Noth et al., 1985)). The value of LLRs lies not in their diagnostic utility but in their ability to shed light on physiological mechanisms in healthy subjects and in patients with motor disorders.

3. Brain stem reflexes, including startle

Brainstem nuclei and circuits are engaged in many activities such as motor preparation, sensory gating, swallowing, breathing, sleep, etc. In humans, neurophysiology helps understanding the circuits involved in such basic functions through, basically, the study of brainstem reflexes. When dealing with brainstem reflexes, the examiner must be aware that response latency and size reflect not only the conduction of impulses along a given reflex circuit, but also the excitability of such circuit, which is under control by supranuclear centers and modulated by a variety of inputs (Valls-Sole, 2012).

Normative data on the most frequently used brainstem reflexes have been published by many authors (Aramideh and Ongerboer de Visser, 2002; Cruccu et al., 2005; Kimura, 1982), which can be used for clinical purposes, but, as in almost all neurophysiological tests, it is convenient for the examiners to gather normative values in the context of their own clinical practice. Among the cranial nerve elicited brainstem reflexes, the most frequently used is the blink reflex to electrical stimulation of trigeminal nerve branches, but other reflexes generated by auditory, visual or somatosensory stimuli can certainly be more informative in specific conditions. Most stimuli impinging on the brainstem, either natural or experimental, induce not only a local reflex reaction, but may lead to effects beyond the specific response recorded, due to activation of the brainstem reticular formation. The typical motor response generated in the reticular formation is the startle reflex (Brown et al., 1991b; Kofler et al., 2001a, 2001b; Landis and Hunt, 1939). This is an involuntary response, which full-blown expression entails a generalized body movement. Less apparent manifestations can be evident in neurophysiologic recordings from specific muscles even when stimuli are intended for the study of other reflex reactions. In fact, there is a tight relationship between the blink and the startle reflex, as depicted in the schematic circuitry of Fig. 5. While brainstem reflexes are robust responses, their circuits incorporate also inhibitory control mechanisms that are called into action to limit inappropriate reflex reactions. One of the most important inhibitory systems is prepulse inhibition, i.e., the inhibition of a reflex response induced by a pre-stimulus that is unable to produce a response by itself (Graham, 1975; Valls-Sole et al., 1999b).

The following paragraphs review the physiology of the blink reflex, the startle reaction and prepulse inhibition, with a touch on other brainstem reflexes with clinical utility.

3.1. Blink reflexes

The blink reflex is easily obtained in clinical settings by just tapping with our fingers on the subject's glabella, as neurologists do in standard physical exams to check cranial nerve function. If
Polyneuropathies may affect the R1 and the R2 to a similar extent (Eekhof et al., 2000; Kimura et al., 1969; Valls-Sole, 2012). The difference in thickness of the arrows pointing to R1 and R2 responses mark their different behavior to conditioning experiments (see text). The R2 response can also be generated by inputs from other sources. Output projections from RF go to motoneurons of the facial and other brainstem nuclei, as well as to spinal cord alpha motoneurons. Additionally, inputs are processed in the prepulse circuit (pP), causing inhibition of the R2 and the startle reflex responses (broken line arrows).

Responses are recorded from the orbicularis oculi muscle in the lower eyelid with surface electromyography electrodes to discrete controlled stimuli, the examiner can also get information on response latency and size. Quantifying the responses is, in fact, the main aim of electrodiagnostic testing. Electrical shocks are well suited for this purpose, as the stimuli are brief.

3.1.1. Blink reflexes to supraorbital stimuli

The blink reflex to electrical stimuli of the supraorbital nerve is the most frequently used brainstem reflex in clinical practice. It differs, though, from the blink reflexes elicited by other forms of stimulation for the presence of the R1 response, an oligosynaptic response mediated by the principal nucleus of the trigeminal nerve at the rostral pons, which appears at a latency of 10–12 ms in the ipsilateral side of the stimulus. The R1 is not concerned with blinking but is of paramount importance for clinical electrodiagnostic assessment of peripheral nerve or brainstem lesions (Kimura, 1982; Marx et al., 2001). The orbicularis oculi response accompanied by blinking is the R2 (Evinger et al., 1984), which appears at a latency around 35 ms, not only in the side where the supraorbital nerve stimulus is applied, but also in the contralateral side (R2c). The R2 and R2c responses are analogous to the blink responses elicited by other forms of stimulation when latency differences due to different fibers conduction velocity and synaptic strength are accounted for (Fig. 6).

In clinical practice, the analysis of the blink reflex allows for distinguishing various patterns of abnormalities (Aramideh et al., 1997; Esteban, 1999; Valls-Sole, 2019): Focal lesions should cause an afferent pattern if they involve the trigeminal nerve, i.e., absent or abnormally delayed responses in both sides to stimuli applied in the affected side, with preservation of those generated by stimuli applied to the unaffected side, or an efferent pattern if they involve the facial nerve, i.e., absent or abnormally delayed responses in the affected side with preservation of those recorded from the unaffected side (Eekhof et al., 2000; Kimura et al., 1969; Valls-Sole, 2013). Polyneuropathies may affect the R1 and the R2 to a similar extent, with the proviso that since the R2 response is mediated by substantially more synapses than the R1, its latency will be less dependent on conduction than on synaptic delay. In contrast, lesions altering the trigeminal neural complex at the brainstem usually affect either the R1 or the R2 responses independently (Cruccu et al., 2005; Kimura and Lyon, 1972). Upper brainstem lesions may preserve the blink reflex to trigeminal nerve stimulation while the blink reflex to median nerve stimulation appears selectively impaired (Leon et al., 2011; Valls-Sole et al., 1997). In contrast, lesions involving the lower brainstem affect the R2 to supraorbital nerve stimulation but preserve the response to median nerve stimulation. Finally, lesions in distant sites or changes due to regeneration after facial nerve lesions may modify the reflex circuit excitability. In case of unilateral lesions, such effect on reflex excitability will be asymmetric and may become apparent when calculating the R2c/R2 size ratio (Cabib et al., 2014; Manca et al., 2000).

The excitability of the blink reflex circuit is under control of supranuclear structures, including the basal ganglia (Basso and Evinger, 1996; Basso et al., 1996) and sensorimotor cortical areas (Berardelli et al., 1983; Fisher et al., 1979). A popular method to measure excitability of the blink reflex circuit is by examining the responses to a test stimulus after the passage of activity induced by a conditioning stimulus, the so-called paired-shock technique (Kimura, 1973). Usually, this procedure shows a time-dependent post-activation decrease of excitability (Valls-Sole, 2012), which is more manifest in the R2 than in the R1 because of more synapses involved. After eliciting the R2 to a supraorbital nerve stimulus in healthy subjects, it takes more than 1 s for another stimulus of the same intensity to induce an R2 response of similar size. The blink reflex excitability recovery curve can be plotted by representing the percentage recovery of the R2 to the second stimulus with respect to the first as a function of time. It is of great clinical utility as it is altered in many neurological disorders (Aktekin et al., 2001; Berardelli et al., 1985; Kimura, 1973; Kumru et al., 2010; Nakashima et al., 1990; Schwingenschuh et al., 2011). Another form of conditioning the blink reflex response is by way of using a prepulse stimulus (Boelhouwer et al., 1991;
sory startle (Alvarez-Blanco et al., 2009) and to assess changes in
1997), it has been used to characterize the so-called somatosen-
(Benbir and Kiziltan, 2014; Miwa et al., 1998; Valls-Sole et al.,
reflex (HBR). Apart from some potential clinical applications
the peripersonal space (Sambo et al., 2012). It is interesting to note
the differences between the two reflexes. 2) Auditory stimuli.
when the trigemino-facial reflex is altered, which shows some of
that the HBR may be preserved in some lower brainstem lesions,
unpredictable reflex cannot be obtained after electrical stimula-

to acquisition of a new eye closing response that appears just
before the puff of air for the subject (or the experimentation ani-

Fig. 7. Startle reaction to an unexpected loud auditory stimulus (110 dB). In A, responses were recorded while the subject was at rest. In B, responses were recorded when the same subject was ready to perform a fast wrist extension movement at perception of the auditory signal. Note the enhancement of the responses recorded from the Orbicularis Oculi (OOc) and the sternocleidomastoid (SCM), and the burst of activity recorded from the wrist extensor muscles (WE), leading to the wrist extension movement (MOV).

Rossi and Scarpini, 1992; Valls-Sole et al., 1999b). The accepted
mechanism by which the R2 is inhibited by a prepulse differ from the one suggested for the paired shock technique, as the prepulse stimulus does not generate a response by its own. Prepulse inhibition is also present for the startle reaction and it will be reviewed below.

3.1.2. Blink reflexes to stimuli other than electrical shocks to the supraorbital nerve

The blink reflex is a ubiquitous response to any stimulus type. Two types of eliciting stimuli other than the electrical stimulation of the supraorbital nerve are of particular interest for physiology and clinical applications: 1) Stimuli to the median nerve at the wrist. This was initially described as the somatosensory blink reflex (Miwa et al., 1998) and is also known today as the hand blink reflex (HBR). Apart from some potential clinical applications (Benbir and Kiziltan, 2014; Miwa et al., 1998; Valls-Sole et al., 1997), it has been used to characterize the so-called somatosensory startle (Alvarez-Blanco et al., 2009) and to assess changes in blink reflex circuit excitability when stimuli are applied within the peripersonal space (Sambo et al., 2012). It is interesting to note that the HBR may be preserved in some lower brainstem lesions, when the trigemino-facial reflex is altered, which shows some of the differences between the two reflexes. 2) Auditory stimuli. Any unexpected sound is sufficient to induce a response of the orbicularis oculi. In fact, auditory stimuli of large intensity are known to induce the startle reaction, which most conspicuous response is closure of the eyes. A circuit has been defined for the auditory blink reflex separate from the circuit of the startle re-

3.2. Other cranial nerve reflexes useful in clinical practice

The mandibular reflex or jaw jerk is of paramount importance in the evaluation of suspected upper brainstem lesions (Hopf, 1994; Marx et al., 2001). The mandibular reflex circuit is particular for the location of the cell bodies. In contrast to those of all other muscles in the body, the proprioceptive neurons of the masseter muscles lie within the neuraxis, protected by the blood–brain barrier from peripheral circulating agents (Graus et al., 1987). This is actually an important piece of information for the diagnosis of some disorders involving immunological aggressions to sensory neurons of the Gasserian ganglia (Valls-Sole et al., 1990).

A silent period follows the excitatory reflex response of the masseter muscles because of after-hyperpolarization inhibition and refractoriness of activated neurons. However, a more reproducable inhibitory reflex can be obtained after electrical stimulation of the mentalis nerve (Cruccu et al., 1989; Ongerboer de Visser et al., 1990). This has interesting clinical applications (Cruccu and Deuschl, 2000) and has been the subject of many studies on trigeminal pain (Wang et al., 1999).

In addition to the electrical methods noted above, a puff of air directed to the cornea also induces a blink reflex. This is the method used to examine classical Pavlovian conditioning circuits of the cerebellum (Solomon et al., 1989). In this method, pairing of low intensity auditory tones with puffs of air to the cornea leads to acquisition of a new eye closing response that appears just before the puff of air for the subject (or the experimentation animal) to prevent the nuisance of corneal stimulation. A similar effect can be observed using electrical methods to evoke the blink reflex. This kind of learning has been attributed to cerebellar circuits (Dimitrova et al., 2002) and has been found abnormal in many disorders including some forms of dementia (Woodruff-Pak, 2001), parkinsonisms (von Lewinski et al., 2013), major depression (Greer et al., 2005) and dystonia (Janssen et al., 2014).

3.3. The startle reflex

The startle reflex is a generalized motor reaction triggered by a stimulus causing surprise or alarm, characterized by a sudden involuntary movement of the body (Brown et al., 1991a; Davis et al., 1982; Dreissen and Tijssen, 2012; Landis and Hunt, 1939; Wilkins et al., 1986). It is a common response of many animal species, which indicates that the motor system conveying the startle reflex must be phylogenetically old. This has been identified as the reticular formation and its descending tracts, mainly the med-

viscerals and the responses typically recorded are from the orbicularis oculi, two different circuits contribute to the response, with either partially superimposed or consecutive activity: the auditory blink reflex and the orbicularis oculi component of the startle reflex (Brown et al., 1991b). The circuit involved in the auditory

3051
The startle reflex is very simple, as direct connections have been identified between the cochlear nuclei and the neurons of the nucleus reticularis pontis caudalis (Davis et al., 1996; Koch, 1999; Nodal and Lopez, 2003), but these neurons are not modality specific and, therefore, the startle reflex can be generated by other inputs such as, among others, high intensity sensory stimuli (Alvarez-Blanco et al., 2009).

Humans and domesticated animals have learnt to control the involuntary motor reactions induced by startling stimuli, although exaggerated reactions can be seen in certain neurological disorders (Tijssen et al., 2002). In fact, in many body movements, we use the reflex activation of the reticulospinal tract incorporated or summed to our voluntary actions, at least concerning movement preparation, anticipatory postural adjustments or automatic movements. An interesting phenomenon occurs when the startled subject is in a state of preparation for execution of a voluntary motor act (Fig. 7). In this circumstance, reaction time will be speeded up and the startle reflex enhanced, a phenomenon known as StartReact (Carlsen et al., 2004; Valls-Sole et al., 1999a).

### 3.4. Prepulse inhibition

Involuntary reflex reactions, such as those induced by a startle, require some form of inhibitory control to adapt to the environment. There probably are many sources of inhibitory inputs to the nucleus reticularis pontis caudalis, but one of the brainstem circuits best studied in animals and humans is prepulse inhibition, defined as the inhibition observed on a reflex response when the reflex-eliciting stimulus is preceded by a weak stimulus that does not induce a response by itself (Blumenthal and Levey, 1989; Graham, 1975; Hoffman and Ison, 1980; Ison and Hoffman, 1983; Swerdlow et al., 1993; Valls-Sole et al., 1999b). Prepulse inhibition was first known as a reflex modification of the effects of weak sensory pre-stimulation and it is now considered an expression of sensory gating (Garcia-Rill et al., 2019). The effects are mediated through the pedunculopontine tegmental nucleus (Koch, 1999).

Although the cholinergic neurons of the pedunculopontine nucleus have long been identified as the ones responsible for inhibition of activity in the nucleus reticularis pontis caudalis, recent evidence suggests that this is not the case (Azzopardi et al., 2018; Fulcher et al., 2020; Garcia-Rill et al., 2019).

The effects of a prepulse are not limited to inhibition but entail some facilitation. This was already observed in early works (Davis et al., 1982; Hoffman and Ison, 1980), but was made evident when investigating the effects of sensory pre-stimulation on the blink reflex to supraorbital nerve stimulation, as there was enhancement of the R1 when the R2 and R2c were inhibited (Ison et al., 1990).

Fig. 8 shows the effects of an auditory prepulse on the blink reflex and the startle reflex. Importantly, the effect of a prepulse on the blink reflex demonstrates not only the inhibition of the response conveyed through the reticular formation (i.e., the R2 and R2c responses), but, also, the early facilitation that precedes inhibition. Inputs from any stimulus, included those generated by weak pre-pulses, reach the brainstem, where it may widely distribute and prime targeted neurons. Even if these neurons are not brought to firing level, they may be readier to fire at the arrival of another facilitatory input. This is the case with facilitation of the R1 response to the supraorbital nerve stimulus. Such window of facilitation depends on the timing of arrival of sensory inputs to the brainstem (Valls-Sole et al., 1999b) and is terminated by a longer lasting inhibition. Prepulse inhibition has been also described for other brainstem reflexes (Gomez-Wong and Valls-Sole, 1996; E. Kiziltan et al., 2019).

Prepulse inhibition may be constantly working because of the many inputs we receive in our daily activities. In fact, many inputs probably interact at the level of the pedunculopontine tegmental nucleus to constantly modulate its outcome for the contextual control of the startle reaction. It turns out that the percentage inhibition caused by a prepulse might differ according to contextual factors. One example of this is the dependence of the prepulse effect on posture and site of sensory stimulation (Versace et al., 2019). There are many aspects of prepulse inhibition left to study.
in healthy subjects and patients. Phenomenologically, failure of prepulse inhibition has been observed in patients with schizophrenia and other psychotic disorders, where it provides for an objective marker of dysfunction of control mechanisms over sensory stimuli (Braff and Geyer, 1990; Fulcher et al., 2020; Li et al., 2020; San-Martin et al., 2020).

3.5. Summary

In summary, the study of brainstem circuits in humans is an important area of neurophysiology, as it may show the integration of sensory inputs on functional motor responses with less complex processing than in cerebral centers. Neurophysiological study of brainstem reflexes brings up two important quantifiable pieces of information in the studies of brainstem circuits: the time it takes for impulses to run through them (which expression is, typically, response latency), and how likely it is to get a response after a controlled stimulus (which is expressed in the size of the response and various excitability tests). Out of the many brainstem reflexes available to neurophysiological study, the blink and startle reflexes, together with their inhibitory control by prepulses, are doubtless the most used in the clinical and research areas. The findings in these tests in healthy subjects and patients allow for interpretation of functional derangements taking place in the brainstem in many neurological disorders.

4. Transcranial magnetic stimulation (TMS and rTMS) investigation of cortical circuits

Transcranial magnetic stimulation is a simple technique for electrical stimulation of the brain through the intact scalp. It uses a time-varying magnetic field to induce an electric current in the brain which is of a very similar shape and duration as that produced by a conventional peripheral nerve stimulator. All TMS pulses are charge-balanced: there is no active anode or cathode and so there can be no net current in the brain. The simplest pulse would therefore be a single sine wave in which the positive and negative phases of the pulse cancel each other. Both positive and negative phases can potentially activate neurons. However, most conventional stimulators produce what is termed a “monophasic” pulse in which the initial pulse in one direction is short-lasting (usually about 100 µs), and followed by a longer lasting pulse in the opposite direction. The first part of the pulse has a high amplitude and activates the brain; the amplitude of the second part is much lower and is below the threshold for neural activation.

Because the magnetic field falls off rapidly with distance from the coil, stimulation is confined to surface structures (e.g. cerebral cortex) unless special designs of coil are used. TMS is not very focal; the usual figure-of-eight coils activate a region about 1–2 cm², but the precise value depends on the intensity of stimulation and the design/size of the coil. More details can be found in the IFCN guidelines for TMS (Pierrot-Deseilligny and Burke, 2012; Rossini et al., 2015).

TMS can be used to stimulate any part of the brain but has most commonly been used to stimulate motor cortex. This is because a single pulse evokes contraction of contralateral muscles that can easily be recorded with EMG (a motor evoked potential or MEP). It gives a direct read-out of the effect of each pulse. This means that when used with a focal coil, it can be used to map out the somatotopy of the motor cortex, at least within the upper limb to investigate changes caused by pathology such as amputation (Gunduz et al., 2020) or tumor growth (LeLafuente and Picht, 2016). A crucial point about TMS of motor cortex (and probably other areas) is that the lowest threshold sites of activation are synaptic terminals in cortical grey matter (Rossini et al., 2015). This means it can tell us something about the excitability of connections within the cortex, and many techniques exploit this fact to probe details of human cortical physiology.

4.1. Single pulse TMS

The simplest TMS method is single pulse TMS of motor cortex in order to evoke an MEP. The pulse activates excitatory synaptic inputs to corticospinal neurones. When these discharge, action potentials are conducted down the corticospinal tract to spinal motoneurons and thence to muscle. The high intrinsic connectivity within the cortex means that synaptic activity continues for several milliseconds after the pulse is applied. The result of this is that the initial synaptic input is followed by several additional inputs that occur regularly at approximately 1.5 ms intervals for 5–10 ms. These cause multiple discharges in the corticospinal tract, known as “indirect” waves, or I-waves, labelled in order of appearance, I1, I2, I3 etc (Patton and Amassian, 1954). The synaptic inputs to corticospinal neurones that are responsible for initiating these waves of descending activity are known as I-wave inputs (Fig. 9). The I1 input to corticospinal neurones is thought to be monosynaptic and the corticospinal activity it produces is responsible for the initial onset of the MEP. When I-waves reach spinal cord they release EPSPs which summate at the motoneurons and determine the amplitude of the MEP.

The most reliable outcome measures are the latency of the MEP and threshold TMS intensity required to evoke a minimal response (Brown et al., 2017). It is important to recall that the latency measure involves the time taken for synaptic activation of corticospinal neurones within cortex, conduction down the corticospinal tract, synaptic transmission to spinal motoneurons, and conduction time to muscle. Nevertheless, in conjunction with measures of the peripheral conduction time from spinal ventral horn to muscle and estimates of synaptic transmission time, the latency can be used to estimate the central conduction time and conduction velocity within corticospinal tract. It is a clinically validated measure useful in diagnosis of MS and cervical spondylosis.

The threshold of the response depends on several factors including the physical distance from the cortex to the surface of the scalp. However, the biologically important factors are the excitability of axonal membrane at the point of activation in cortex and the excitability of synaptic connections in brain and spinal cord. As such, threshold is affected by drugs that act on voltage-gated sodium channels such as carbamazepine or on glutamatergic transmission, such as ketamine (Ziemann et al., 2015). An important point is that threshold is also influenced by ongoing levels of activity in the pathway since these will affect efficiency of transmission in the synaptic connections. Thus, threshold at rest is higher than during activity because the post-synaptic neurones are more easily discharged if their membrane potential is depolarized by ongoing synaptic input.

An unexpected property of motor cortex TMS is that it is directionally selective (Day et al., 1989) (Fig. 9). A standard figure-of-eight coil induces a directional current flowing along the intersection of the two “wings”. When a monophasic pulse is used, the threshold of activating hand muscles is lowest if current is induced in an approximately posterior-anterior direction perpendicular to the line of the central sulcus. This direction of current preferentially excites I1 inputs to corticospinal neurones. If the current is reversed to anterior-posterior, it tends to excite I3 inputs, resulting in a later onset of the MEP. Rotating the coil by 90 degrees so that it is approximately parallel to the central sulcus can often result is direct activation of corticospinal tract axons. This is known as “direct” activation (as opposed to synaptic activation) and leads to shortest latency MEP (Di Lazzaro, 2013).
To summarise, because a single TMS pulse has a very high temporal resolution and can probe how cortical excitability of different synaptic inputs to corticospinal neurones changes over time, it can be used for example, at high resolution during the course of a movement or over long periods such as during motor learning. In addition, because an MEP represents the summed output of a specific set of corticospinal neurones destined for that muscle, it provides high spatial resolution of motor cortical physiology. For example, simultaneous recording from several muscles has been used to document "surround inhibition" of relaxed muscles during focal contraction of a neighbouring muscle (Sohn and Hallett, 2004).

Although the techniques described so far give impressive detail about the excitability of neural circuits that produce activity in the population of corticospinal fibres that produces the MEP, they represent only a small fraction of the total neural circuitry in the cortex. Probing these requires more complex methods.

4.1.1. Probing connectivity within the motor cortex

Two main techniques probe inhibitory circuits in motor cortex. A single TMS pulse to an actively contracting muscle produces an MEP that is followed by a silent period that interrupts ongoing EMG activity. The initial part of this EMG silence is due to inhibition in the spinal cord resulting from (a) the refractory period of motoneurones that discharged in the MEP, (b) Renshaw inhibition, and (c) activation of other local inhibitory circuits such as la inhibitory interneurons and propriospinal inhibitory neurones. However, the later part of the silent period from about 75 ms onwards is due to reduced excitability in motor cortex that is thought to result from activity in both GABA-A and GABA-B circuits. The duration of the silent period increases with stimulus intensity in a sigmoid fashion and plateaus around 200–300 ms.

The silent period represents a pause or disruption in ongoing volitional motor cortex activity, and disrupts performance of any task that is being performed at the time of the stimulus. This has been termed a "virtual lesion". A similar effect is thought to occur with stimulation of any cortical area: ongoing processing is disrupted for up to 100–200 ms by a TMS pulse (Walsh and Cowey, 2000). This can be used very effectively to probe whether and when a particular cortical area participates in a task. For example, a TMS pulse to primary visual cortex given 100 ms after presentation of a brief visual stimulus abolishes perception of the stimulus, effectively creating a short-lasting scotoma in the visual field (Maccabee et al., 1991).

Inhibition can also be probed in more detail with paired pulse methods in which two TMS pulses are applied through the same coil. In short-interval intracortical inhibition (SICI), an initial submotor threshold (conditioning) pulse precedes a suprathreshold second (test) pulse. If the interstimulus interval (ISI) is approximately 2–5 ms, then the MEP evoked by the test pulse is smaller.

![Fig. 9. Descending volleys evoked by electrical and magnetic stimulation and by paired pulse magnetic stimulation. Each trace is the average of the responses to 10–25 cortical stimuli, recording shown in the three columns have been obtained in three different subjects. Electrical anodal stimulation at threshold intensity evokes the earliest volley that is termed D-wave. Low intensity magnetic stimulation with a posterior-anterior (PA) induced current in the brain evokes a single descending wave with a latency about 1 ms longer than the D-wave evoked by electrical stimulation that is termed I1 wave. At intermediate intensity later I-waves are evoked and at high intensity, an earlier small wave with the same latency of the D wave evoked by electrical stimulation appears. Magnetic stimulation with a latero-medial (LM) induced current in the brain preferentially evokes D-wave activity. With biphasic magnetic stimulation the earliest volley has a latency of about 0.4 millisecond longer than the D wave evoked by LM magnetic stimulation. Because of its longer latency, it is suggested that the D wave evoked by biphasic stimulation is initiated closer to the cell body of the PTNs than the conventional D wave evoked by LM magnetic stimulation and anodal stimulation and it is termed “proximal D wave”). On the right, epidural volleys evoked by test magnetic stimulus alone (solid trace) and by test magnetic stimulus preceded by a subthreshold conditioning stimulus at 3 milliseconds interstimulus interval (dotted trace). The test stimulus evokes multiple descending waves. There is a clear suppression of the late corticospinal volley when the test magnetic stimulus is preceded by the subthreshold conditioning stimulus. From (Di Lazzaro and Rothwell, 2014) with permission.](image-url)
than if the test pulse was applied alone (Kujirai et al., 1993). Drug studies have shown that the effect is due to activity in a GABA-A connection (Ziemann et al., 2015). The depth of inhibition depends on the intensity of the conditioning pulse in a "U"-shaped fashion, reaching a maximum around 80–90% resting motor threshold and becoming facilitatory at higher intensities. Several lines of evidence show that SICI suppresses late I-wave inputs (I3 and onwards) to corticospinal neurones, but has little effect on I1 and I2 inputs (Di Lazzaro and Rothwell, 2014). Because of this, SICI is affected by both the intensity and the orientation of the test pulse, both of which alter the proportion of early and late I-wave contributions to the MEP. SICI is also less effective when examined in actively contracting muscle for the same reasons. Finally, note that SICI can also be observed with a very short ISI of 1 ms, but its mechanism is less clear. There is probably a period of axonal refractoriness following the conditioning pulse (Fisher et al., 2002).

As with MEPs, SICI has a high temporal resolution and can probe the excitability of GABA-A inhibition at different times during performance of a task. For example, SICI is reduced just before the start of a voluntary contraction in the same way as the brake on a car is removed just before moving off (Reynolds and Ashby, 1999).

If the ISI is increased to 10–20 ms, test MEPs are facilitated (intracortical facilitation, ICF). The threshold for this effect is slightly greater than that for SICI, suggesting that it is a separate phenomenon (Kujirai et al., 1993). Intracortical facilitation has been proposed to be a glutamatergic effect that overlies continuing GABA-A inhibition. However, there is some question as to whether it may also be contaminated with a subtle spinal effect (Wiegel et al., 2018).

Longer-lasting GABA-B inhibition is probed in a similar way, but in this case, the threshold is higher than for GABA-A effects. In long-interval intracortical inhibition (LICI), the conditioning and test pulses are both suprathreshold and the ISI is >50 ms. The depth and duration of the inhibition depend on the intensity of the conditioning stimulus, and it is followed by a shorter period of late cortical disinhibition (Cash et al., 2010; Rossini et al., 2015).

One final variation on the paired pulse design is short-interval intracortical facilitation (SICF) in which a just-suprathreshold pulse is followed by a just-subthreshold pulse at intervals from 1–4 or 5 ms. MEP facilitation occurs at a series of ISIs around 1.1–1.5 ms, 2.3–2.9 ms and 4.1–4.4 ms, interrupted by periods in which the MEP is unaffected. The hypothesis is that this probes the generation of I-waves since the SICF peaks occur at the same intervals as I-waves (Peurala et al., 2008; Tokimura et al., 1996; Ziemann, 2020).

4.1.2. Probing connectivity to motor cortex

Paired pulse designs are easily expanded to probe inputs to motor cortex from other parts of the CNS (e.g. Fig. 10). As with the MEP and SICI, the effects change during and in preparation for different tasks. For example, a single conventional electrical stimulus to the median or ulnar nerves can suppress responses evoked in hand muscles by a TMS pulse applied about 20 ms later. This is known as short-latency afferent inhibition (SAI) (Tokimura et al., 2000): afferent input from the peripheral stimulus arrives indirectly at motor cortex where it inhibits production of the MEP. The effect increases with intensity of the conditioning stimulus and its depth is modulated by cholinergic drugs, suggesting that it may be a useful monitor of cholinergic function (Ziemann et al., 2015). Like SICI, SAI appears to be reduced prior to finger movement (Cho et al., 2016), and seems to preferentially target late I-wave inputs to corticospinal neurones (Tokimura et al., 2000). SAI is followed by a later and less well studied period of inhibition, known as long-interval afferent inhibition (LAI) (Chen et al., 1999).

If two TMS coils are available, then it is possible to probe inputs to the motor cortex from other parts of the cerebral cortex and cerebellum (Perbert et al., 1992). Inter-hemispheric inhibition (IHI) describes how stimulation of the motor cortex of one hemi-
sphere suppresses MEPs evoked from the opposite hemisphere approximately 8 ms later. It is thought that the first stimulus activates transcallosal neurones in layer 3 of cortex that connect to inhibitory interneurons in the opposite hemisphere to suppress MEPs. Thus, the onset of inhibition is a measure of transcallosal conduction time. The duration and depth of inhibition depend on the conditioning stimulus intensity. With care it is also possible to observe a period of transient facilitation just prior to the onset of IHI, which may represent an initial facilitation by transcallosal fibres (Hanajima et al., 2001).

Inhibitory and facilitatory effects on motor cortex have been described from a large number of different cortical sites in frontal and parietal areas, such as premotor or supplementary motor cortex, or posterior parietal regions (e.g. (Arai et al., 2012; Civardi et al., 2001; Koch et al., 2008; Koch et al., 2007)). Use of larger coils that activate deeper structures has also shown effects from cerebellum (Ugawa et al., 1995). The excitability of all these inputs changes with task, and careful probing can reveal how and when specific behaviours involve activity in specific sets of cortico-cortical connections. It is a powerful tool to explore dynamic connectivity in the brain in both health and disease. Finally, it should be pointed out that many of these connections interact with each other; for example, SAI affects SICI, SICI affects LICI etc (Chen, 2004). Probing these interactions can sometimes involve up to three TMS pulses (triple stimulation).

4.2. Repetitive TMS and synaptic plasticity

The first TMS machines were only capable of generating one pulse every 3–4 s. However, it is now possible to deliver repeated pulses of TMS at frequencies up to 50 or 100 Hz. The main technical limitation is coil heating particularly at high frequencies and high intensities; the main practical limitation is safety since repeated stimuli can provoke seizures even in healthy individuals. Safety guidelines are available to prevent this (Rossi et al., 2021). One important feature to note is that most rTMS machines produce biphasic pulses rather than the “unidirectional” pulses produced by most single pulse machines. This is because biphasic pulses can recycle energy back into the capacitor of the stimulator, reducing the time taken for it to charge up for the next pulse. It means that stimulus direction is much less important in rTMS unless special devices are used that can generate repetitive unidirectional pulses.
If the position of the TMS coil is constant, then repeated stimuli will activate the same synapses repeatedly. In animal experiments, this can lead to changes synaptic effectiveness known as long-term depression or long-term potentiation (Bliss and Lomo, 1973). In the motor cortex, rTMS seems to produce analogous effects: application of 1000 stimuli at 1 Hz can reduce the excitability of motor cortex for up to 30 min (i.e., single pulse MEPs are smaller than before rTMS) (Chen et al., 1997) whereas higher frequencies such as 5 Hz or more can increase excitability for about the same length of time (Huang et al., 2017; Peinemann et al., 2004). Since spinal H-reflexes are unaffected the effect is thought to occur within the motor cortex. The effects are variable both within and between individuals and a number of different protocols have been proposed to increase the effects (Huang et al., 2017). The protocols differ in the stimulus intensity, number of pulses, and the rate and regularity of stimulation (see Fig. 11). Unsurprisingly, there are a large number of possible combinations. However the main ones in use at present and for which we have most data are regular rTMS (at 1 Hz (inhibitory) or 5–20 Hz (excitatory)), theta burst stimulation (TBS) (Huang et al., 2005), and quadrupulse stimulation (QPS) (Hamada and Ugawa, 2010). Note that higher frequency rTMS is usually applied for just 1–2 s followed by a longer pause before repeating until the desired number of pulses is applied. Stimulation intensity is usually around motor threshold or above.

In theta burst stimulation, pulses are delivered in bursts of 3–5 pulses at high frequency (e.g., 50 Hz) that are repeated at 5 Hz (Huang et al., 2005). Given the involvement of high frequencies, for safety reasons stimulation intensity is usually below motor threshold (usually 80% active motor threshold), although higher stimulus intensities have been used safely in the treatment of depression (Bakker et al., 2015). The advantage of TBS is that a large number of pulses can be delivered in a short time: continuous TBS (cTBS) delivers 600 pulses in only 40 s. It has an inhibitory effect lasting about 30 min. Intermittent TBS (iTBS) arranges the pulses in a different way: TBS is delivered for 2 s, paused for 8 s and then repeated. Six hundred pulses can be applied in 102 s and has an excitatory effect. Quadrupulse stimulation differs from TBS and rTMS in that it uses monophasic pulses from 4 separate stimulators delivered through the same coil (Hamada and Ugawa, 2010). Four pulses are applied with either short ISIs (1–5 ms) or long ISIs (30–100 ms), and then repeated every 5 s for 30 min using a subthreshold stimulus intensity (90% active motor threshold). Quadrupulse with ISI = 5 ms facilitates motor cortex whereas ISI = 50 ms suppresses excitability.

The repetitive TMS methods above depend on changes in postsynaptic calcium levels that are brought about by repetitive activation of the same sets of synapses (Ziemann et al., 2008). However, TMS methods can also induce synaptic plasticity through Hebbian pairing of two sets of input to the same output neurones. A classic example is somatosensory paired associative stimulation (PAS) in which electrical stimulation of median or ulnar nerves is paired once every 5 or 10 s with a TMS pulse to motor cortex about 25 ms later (Stefan et al., 2000). This is effectively repetitive SAI and usually employs 100–200 stimulus pairs. Given that SAI is inhibitory it is perhaps surprising that this version of PAS increases cortical excitability. If the interval between peripheral and central stimulation is 10 ms then the effect is inhibitory. It is thought that repeated pairings of the inputs cause one or other of the inputs to be strengthened (Stefan et al., 2002; Weise et al., 2013). The rules determining which synapses change effectiveness vary between systems and depend on the order and interval between the inputs (spike-timing dependent plasticity). This is why PAS25 has an excitatory effect whereas PAS10 is inhibitory. The same approach can be used with connections within the brain: for example, repeatedly pairing parietal and motor cortex stimulation can change the effectiveness of parietal inputs.

A final variation on TMS plasticity methods is to study the interaction between two successive applications of the same or different protocols (Karabanov et al., 2015). Logic as well as data from many reduced preparations suggests that it cannot be possible to strengthen a synapse ad infinitum; there must be some limit. In fact, strengthening a synapse a little may be easy but further strengthening becomes successively more difficult until saturation is reached. Simultaneously, if becomes easier and easier to weaken the strengthened synapse. This is called homeostatic plasticity, which the CNS employs to maintain overall activity at a constant level. For example, a short period of PAS25 may increase excitability, but if a second period is applied 30 min later, the effect becomes inhibitory. Conversely if the first block was PAS15, which produces an inhibitory effect, then the effect of the subsequent PAS25 was enhanced facilitation (Muller et al., 2007).

There are many variations on this technique, some of which combine TMS probes of plasticity with plasticity induced by behaviour motor learning (Karabanov et al., 2015). The effect depends on the interval between the plasticity protocols and may be homeostatic, as described above, or “priming” in which case effects summate. Importantly, the effect of the first protocol may be too weak to produce any obvious changes in MEP on its own but still affect the response to a second protocol.

Plasticity methods are not only useful for investigating central neurophysiology and health and disease, but they may also be used therapeutically either when given alone, as in treatment for depression, or in conjunction with a behavioural therapy, as in post-stroke rehabilitation.

5. TMS-EEG

The combination of transcranial magnetic stimulation (TMS) with electroencephalography (EEG) to record TMS responses from the brain (for review, (Ilmoniemi and Kicic, 2010; Tremblay et al., 2019)) is a more recent method compared to recording TMS evoked responses with electromyography (EMG) from muscle, since TMS-EEG required the development of TMS-compatible EEG amplifiers (Ilmoniemi et al., 1997; Virtanen et al., 1999). Moreover, hard- and software solutions had to developed to eliminate the large and long-lasting electromagnetic artifact induced by the TMS coil discharge in the EEG and allow recordings of artifact-free TMS-evoked potentials (TEPs) within a few milliseconds after the TMS pulse (Ilmoniemi et al., 1997; Massimini et al., 2005; Paus et al., 2001). Early TEPs can also be contaminated by large muscle responses due to the direct activation of scalp muscles (Mutanen et al., 2013; Rogasch et al., 2013). Muscle activation can be avoided to a significant extent by coil placement close to the midline, if feasible by experimental design (Massimini et al., 2005). In addition, algorithms have been introduced to efficiently suppress TMS-related muscle artifacts in EEG while retaining the neuronal EEG signals (Mutanen et al., 2016). Late TEPs with latencies >60–70 ms can be contaminated by auditory evoked potentials induced by the TMS click and somatosensory evoked potentials caused by direct excitation of cutaneous afferents in the scalp (Ahn and Fröhlich, 2021; Conde et al., 2019; Rogasch et al., 2014). They can be substantially reduced by applying masking noise through earplugs and use of a spacer between coil and scalp (Massimini et al., 2005). These peripheral evoked potentials are typically midline potentials, that can be differentiated from the lateralized TEPs are largely eliminated by data postprocessing (Rogasch et al., 2014; Wu et al., 2018). They may be controlled by a realistic sham condition that ideally should be indistinguishable from real TMS except the missing TMS pulse, but the optimal solution of a realistic sham is still under development (Belardinelli et al., 2019; Siebner et al., 2019). Moreover, when stimulating motor cor-
It is advisable to use stimulation intensity below motor evoked potential threshold to avoid muscle twitches that would lead to contamination of TEPs by somatosensory reafferent signals. This is possible because TEPs have a much lower threshold compared to motor evoked potentials (Kahkonen et al., 2005; Komssi et al., 2004). Provided that the necessary precautions are taken to avoid, control or eliminate artifacts and peripherally evoked potentials, then TMS-EEG offers several relevant advantages when compared to other techniques, such as TMS-EMG, resting-state EEG or functional MRI: First, it assesses cortical excitability and effective connectivity with high temporal resolution in the order of milliseconds that is proportionate to direct neuronal responses and propagated neural network activity evoked by the TMS pulse; second, TMS-EEG bypasses sensory and motor pathways and, therefore, does not depend on the integrity of sensory and motor systems, allowing assessment even of deafferented or paralyzed patients. A disadvantage is need of obtaining typically at least 100 TMS-EEG trials of good technical and data quality for trial averaging, making this a relatively time consuming procedure.

Single-pulse TMS of the primary motor cortex with the induced current in the brain oriented from posterior to anterior results in TEPs characterized by a sequence of positive and negative deflections occurring at remarkably preserved latencies (for review, Ilmoniemi and Kicic, 2010; Tremblay et al., 2019). The stimulated motor cortex exhibits responses with short latency of 3–7 ms, while the motor cortex in the non-stimulated contralateral hemisphere responds with latencies of 17–28 ms (Ilmoniemi et al., 1997; Komssi et al., 2002). The peak-to-peak amplitude of the N15-P30 complex at the site of stimulation over motor cortex (Fig. 12), i.e., a negativity at 15 ms followed by a positivity at 30 ms, correlates with MEP amplitude (Maki and Ilmoniemi, 2010), is strongly affected by TMS coil orientation (Bonato et al., 2006) and, therefore, represents local excitability of motor cortex and corticospinal tract neurons. These early components of motor cortex TEPs are followed by two highly reproducible, EEG negative deflections peaking at around 45 ms (N45) and 100 ms (N100) (Bonato et al., 2006; Lioumis et al., 2009) (Fig. 12). The N45 is a dipolar potential with the cortical generators of a positivity posterior and the negativity (N45) localized anterior to the stimulation site reaching into the contralateral frontal cortex (Komssi et al., 2004; Litvak et al., 2007; Paus et al., 2001). The N100 is a negativity originating close to the site of motor cortex stimulation (Komssi et al., 2004). Preparation for voluntary hand movement results in reduction of N100 amplitude (Kicic et al., 2008; Nikulin et al., 2003; Yamanaka et al., 2013) and N100 amplitude is significantly correlated with the duration of the cortical silent period (Farzan et al., 2013). These characteristics suggest that the N100 reflects inhibitory neuronal activity. TEP components with even longer latencies (e.g., P180) are not considered here as they probably reflect to significant extent peripherally evoked potentials. Pharmacological profiling has increased further the insight into TEP physiology (for review, Darmani and Ziemann, 2019). The P25 amplitude is decreased by voltage-gated sodium channel blockers (carbamazepine) (Darmani et al., 2019), similar to their decreasing action on excitability of the corticospinal neurons as reflected by decrease in motor evoked potential amplitude. The N45 amplitude is decreased by positive allosteric modulators at GABAA receptors (alprazolam, diazepam, zolpidem) (Premoli et al., 2014) and NMDA receptor antagonists (dextromethorphan) (Belardinelli et al., 2021).
suggesting that the N45 reflects a balance of GABAergic inhibitory and glutamatergic excitatory neural activity. The P60 (P70) potential is decreased by AMPA receptor antagonists (perampanel), providing evidence for this potential to represent fast ionotropic glutamatergic neural activity (Belardinelli et al., 2021). The N100 amplitude at the site of motor cortex stimulation is increased by GABAB receptor agonists (baclofen) (Premoli et al., 2014), corroborating the view that this potential reflects long-lasting GABABergic cortical inhibition.

TEPs exhibit good test–retest reliability (Casarotto et al., 2010; Lioumis et al., 2009) making them applicable for longitudinal studies or assessment of therapeutic interventions in disease. TEPs obtained by motor cortex stimulation are sensitive to experimental manipulation such as variation of stimulus intensity or orientation of the induced current in the brain (Bonato et al., 2006; Komssi et al., 2004). Stimulation of cortical areas other than motor cortex results in a variation of TEP components and frequency content (Rosanova et al., 2009). Finally, brain state may significantly alter TEPs, such as sleep (Massimini et al., 2005), alcohol consumption (Kahkonen et al., 2001), or the phase of the ongoing sensorimotor \( \mu \)-oscillation at the time of the TMS pulse (Desideri et al., 2019).

Investigations of cortical oscillations focus on the study of TMS-induced effects in the frequency domain (Pellicciari et al., 2017; Thut and Miniussi, 2009). In contrast to TEPs, which are time-locked to the TMS pulse, TMS-induced increases in synchronization of EEG oscillations in the beta (15–30 Hz) and alpha frequency range (8–12 Hz) within the first 200 ms are not time-locked to the TMS pulse (Fuggetta et al., 2005; Paus et al., 2001). TMS-induced oscillations are analyzed by a time–frequency approach, such as wavelet decomposition and short-time Fourier transforms, where a sliding window measures TMS-induced oscillatory power across time and frequency (Pellicciari et al., 2017).

TMS-EEG measurements in patients with movement disorders have only started to emerge. For example, the P60 TEP component was increased in patients with Parkinson’s disease at the onset of re-emergent tremor, indicating a role of motor cortex in this phenomenon (Leodori et al., 2020). In another study, TMS-induced oscillations in the beta frequency range were exaggerated in patients with Parkinson’s disease in the unoperated hemisphere but normalized on the side of thalamotomy compared to healthy subjects, providing evidence for hypersynchronized beta oscillations in motor cortex as a pathophysiological correlate of the akinetic-rigid syndrome (Van Der Werf and Paus, 2006).

6. Modulating brain activity (tDCS and tACS)

Non-invasive neuromodulation techniques using electrical currents, such as transcranial direct current stimulation (tDCS) and transcranial alternating current stimulation (tACS) enable researchers and health care professionals to gain unique insight into brain function and to treat symptoms in a number of neurological and psychiatric conditions (Chase et al., 2020). The first version to be developed was tDCS, in which low-intensity direct current (DC) was applied through scalp electrodes; it was followed...
by tACS, in which the DC current was replaced by alternating currents of various frequencies (Fig. 13). Both produce “on-line” and “off-line” effects. The former are those that occur during stimulation, while the latter effects are those that persist after the end of stimulation (also called “after-effects”).

6.1. Basic mechanisms of tDCS

The low intensity currents applied during tDCS (typically 1–2 mA) change the transmembrane potential of cortical neurons: whether neurons are depolarized or hyperpolarized depends on the direction of the electrical current relative to neuronal orientation (Datta et al., 2011; Radman et al., 2009). Consider a large pyramidal neuron of cortical layer V oriented perpendicular to the scalp surface: a surface anode will depolarise the cell body and basal dendrites but hyperpolarise the apical dendrites. In contrast, the same anode will have no effect on a pyramidal neuron lying parallel to the scalp surface (e.g. in the wall of a sulcus). The amount of polarisation depends on the intensity of the applied current and also on the dimensions of the neuron. A large pyramidal neuron will be polarised more than a smaller neuron. This is why the strength of the stimulus is sometimes given in terms of the electrical field it induces in the brain, which is expressed as mV/mm. A field of 1 mV/mm will produce a maximum of 1 mV difference in potential between the tip and base of a 1 mm pyramidal neuron, so long as the neuron is aligned parallel to the direction of the field. The actual depolarisation will always be less than 1 mV depending on the properties of the neuronal membrane, sometimes called the “coupling constant”. In most cases, the maximum levels of polarisation are of the order of 0.2 to 0.5 mV, which is well below the threshold for inducing action potentials. However, polarisation of this level can still affect the ongoing firing rates of neurones and thus excitability at both local and network levels (Opitz et al., 2016; Radman et al., 2009).

Theoretical models as well as experimental data in animal experiments show that tDCS also influences the excitability of synaptic terminals, which can be more effective than that on single neurones (Rahman et al., 2013). Again, the magnitude of the effect depends on the orientation of the synaptic terminals with respect to the applied field. In contrast to polarisation of the soma, where depolarisation increases excitability, at the synapse, hyperpolarisation usually increases excitability (Dudel, 1971). The usual explanation for this is that hyperpolarisation increases the amplitude of presynaptic action potentials, increasing calcium release and the probability of vesicle release.

Finally, it should be noted that the effects of tDCS will also depend on the level of activity in the neurons that are being stimulated. Synaptic activity in a neuron reduces its membrane resistance, and this will change the sensitivity to electric fields produced by tDCS.

The take-home message from such work is that it is difficult to predict what the net effect of tDCS will be, since this depends critically on the orientation of neurones and synapses as well as the level of activity in each. Importantly, it means that anodal tDCS is not necessarily facilitatory and cathodal tDCS inhibitory. In addition, the field between the electrodes may also influence neural activity (Rawji et al., 2018), although the effect will be smaller than directly under the electrodes where the field is spatially more restricted.

6.1.1. tDCS effects on human motor cortex

The effects of DC polarisation of cortex had been studied in animal preparations for many years (Bindman et al., 1962, 1964; Creutzfeldt et al., 1962), and had confirmed that anodal polarisation increases the firing rate of neurones, consistent with membrane depolarisation, whereas the opposite occurs with cathodal polarisation. Studies had also been performed in humans (Costain et al., 1964; Lippold and Redfearn, 1964; Redfearn et al., 1964), but were hampered by the fact that at the time there was no direct measure of cortical excitability. The results were challenged (Lifshitz and Harper, 1968) and the technique fell from favour. It was reinvestigated some 30 years later by Nitsche and Paulus and Priori and colleagues (Nitsche and Paulus, 2000; Priori et al., 1998) who had the advantage that they could use TMS to probe excitability during and after tDCS.

Since then, the majority of studies have reported that when applied over the M1 of healthy individuals, anodal tDCS increases the amplitude of MEPs and conversely, cathodal tDCS decreases them (Nitsche and Paulus, 2000). The effects occur both during and after offset of tDCS. “On-line” effects are thought to be the direct result of depolarisation/hyperpolarisation of cortical pyramidal neurons, whereas the “after-effects” are thought to involve modifications of synaptic plasticity via changes in the regulation of N-methyl-D-aspartate receptors (NMDAR) in a long-term potentiation (LTP)-like or long-term depression (LTD)-like fashion. Experiments with magnetic resonance spectroscopy (MRS) suggest that there may also be effects on levels of GABA transmission that could potentially gate both the anodal and cathodal tDCS-induced NMDAR plasticity (Stagg et al., 2018). The magnitude of the neuronal change depends on many factors, including the type, intensity, duration of stimulation and also on physiological factors, such as the state of the brain, before and during stimulation. In addition to its local effects, tDCS also leads to connectional changes at the neuronal network level, which may influence functional connectivity in both cortical and subcortical networks (Polania et al., 2018).

However, a number of papers have drawn attention to the large variability in response to tDCS both between and within individuals, with several studies failing to observe any net effects (Abdelmoula et al., 2019; Lopez-Alonso et al., 2014; Lopez-Alonso et al., 2015; Vannorsdall et al., 2016; Wiethoff et al., 2014; Wiltshire and Watkins, 2020; Wrightson et al., 2020). The discrepancies probably arise because, as noted above, the results of tDCS depend on so many factors, including the precise orientation neuronal/synaptic populations and current flow, the intensity of current in each individual, and the effects of ongoing (or past) brain states, or individual genetic susceptibility. These are so difficult to control that the predicting the overall effect of tDCS in any individual is highly problematic. tDCS is discussed in further detail in chapter 10, with a focus on therapeutic applications.

6.2. Basic mechanisms of tACS

During transcranial alternating current stimulation (tACS) oscillating electric fields are induced in the brain, by applying low intensity currents that periodically reverse direction via scalp electrodes (Antal et al., 2008). Therefore, when averaged over time the mean membrane potential is not affected by tACS. Nevertheless, the repeated depolarizing and hyperpolarizing effects on the neurones are assumed to be strong enough to modify ongoing neuronal activity in order to induce online and after-effects (Frohlich, 2015).

Stimulation is delivered with constant current devices in order to control for individual differences in scalp resistance. tACS is typically applied in open-loop stimulation, via at least two electrodes. Closed-loop stimulation is also possible by controlling the stimulation parameters using another signal (e.g. by using EEG) and it might offer better neuromodulatory efficacy, however, it is technically more difficult (Lustenberger et al., 2016).

As noted above, the electric fields are usually at their peak under the electrodes, however, depending on the location of the return electrode or at the case of multiple electrodes neuronal net-
The applicability of tACS typically encompasses sinusoidally oscillating current without DC offset at a single frequency (for the most recently used nomenclatures see: (Bikson et al., 2019)). However, several variations have been introduced, including tACS with DC offset (Marshall et al., 2006), or applying a combination of frequencies such as theta-gamma coupling (Alekseichuk et al., 2016b). The possible spectrum of stimulation protocols may be indefinite: for example, the sinusoidal waveform may be biased, biphasic components can vary in amplitude and frequency, or a combination of sinusoids could be used.

Before human experimental applications began (Antal et al., 2008) several animal studies found that the mechanism of tACS is based on entrainment of brain oscillations, as shown for example by modulation of active Purkinje cell activity by AC fields (Chan et al., 1988). Similarly, Francis et al (Francis et al., 2003) demonstrated that electric pulses of 140 μV/mm a peak amplitude were sufficient to increase the firing rate of single neurons in rat hippocampal slices. Nevertheless, the stimulation protocols used in these studies did not resemble the classically, recently used sinusoidal tACS in human studies. Later animal studies using sinusoidal stimulation and low intensities (e.g. (Frohlich and McCormick, 2010; Reato et al., 2010)) revealed a threshold of 0.2–0.5 mV/mm AC fields that were sufficient to modulate the ongoing neural activity so long as there was complete alignment between the direction of applied field and the longitudinal axis of the neuron.

In humans the finite element method (FEM) modeling of the electric fields has shown a significant shunting through the scalp, due to the relatively good conductivity of the skin and low conductivity of the skull. From measurements in primates and in epilepsy patients with implanted electrodes (Huang et al., 2017; Optiz et al., 2016) Liu and colleagues (Liu et al., 2018) estimated that in humans, TES applied at ± 1 mA peak intensity (2 mA peak to peak) generates < 0.5 mV/mm electric fields in the human brain which would be sufficient to generate 0.1–0.2 mV changes in the membrane potential of cells within the stimulated area. Generally, the magnitude of the neural entrainment induced by tACS depends on the difference between the applied stimulation frequency and the frequency of the endogenous neural activity. This implies that in order for tACS to modulate brain oscillations via entrainment, there has to be a pre-existing sustained oscillation before the stimulation starts. tACS at frequencies similar to the endogenous oscillations (the “Eigenfrequency”). At this frequency, it is possible to induce stronger neural entrainment even using low TACS intensities. However, if the stimulation frequency is far from the endogenous oscillation frequency, less or no synchronization occurs. At his case, if the intensity of tACS increases, the synchronization regions will become wider in frequency. Due to its shape, the synchronization region is referred to as an Arnold tongue (Salchow et al., 2016). tACS can also induce effects on frequencies bands that are different from the applied frequency e.g. on the second harmonic frequency or due to cross-frequency coupling properties (see e.g. (Canolty et al., 2006)).

6.2.1. Neurophysiological evidence of tACS in humans – How the effect and after-effect of stimulation can be measured

In human subjects, changes of neural activity/excitability can be measured non-invasively using several methods including measurement of motor evoked potentials (MEPs), electroencephalography (EEG) or magnetoencephalography (MEG). The currents in human tACS studies are typically applied via a bipolar electrode configuration, e.g. when the primary motor cortex is stimulated one electrode is placed over the target area with another over the contralateral orbit. In this case, the most problematic issue is related to the induction of phosphenes during stimulation, due to current flowing via the retina (Paulus, 2010; Schutter and Hortensius, 2010). Multi-site electrode configuration arranged in centre-surround geometry can solve this problem (Datta et al., 2009) because that way the induced field is more focused and the induction of phosphenes can be reduced. Generally, to study the efficacy of tACS is challenging, because a natural consequence of entrainment is that several parameters of the oscillation are manipulated at once. Therefore, different experimental protocols are necessary to test whether a given oscillation’s frequency, and intensity, phase are effective.

The systematic research application of sinusoidal tACS in humans started with Antal and colleagues (Antal et al., 2008), using MEP measurements and was followed by many other studies (e.g.(Bland and Sale, 2019; Grabner et al., 2018; Jones et al., 2019; Kasten and Herrmann, 2019; Ketz et al., 2018; Lorenz et al., 2019)). Most of the human investigations used tACS frequencies in the EEG-detectable range (0.5–70 Hz), because the aim of the studies was to interact oscillations in the EEG range. Nevertheless, tACS in the kHz frequency range can also modulate cortical excitability (Chai et al., 2011).

In the first tACS-EEG combination study, an enhancement of the EEG alpha band amplitude was seen at the posterior part of the brain after 10 Hz tACS for 10 minutes, with after-effects for ~ 30 minutes (Helfrich et al., 2014; Neuling et al., 2013). Measurements are usually done before and after stimulation because the strong artifact that is induced by tACS, renders EEG and MEG recording difficult to analyse and interpret. Recent studies suggested that the stimulation-artefact can be mitigated using spatial filtering (e.g. (Kasten and Herrmann, 2019)).

The effect of tACS can also be measured using blood oxygenation level dependent (BOLD) functional magnetic resonance imaging (fMRI), however, only indirectly. Nevertheless, the results are somewhat contradictory, e.g. it was shown that tACS applied at the individual alpha frequency reduced the amplitude of the BOLD response to visual stimuli (Vosskuhl et al., 2016) but 10 Hz tACS (i.e., not at the individual alpha frequency) showed an effect on BOLD after the end of the stimulation but not during stimulation (Alekseichuk et al., 2016a).

6.2.1.1. Different forms of tACS are also used in research and increasingly in the clinic. Random noise stimulation: The application of random noise (or white noise, i.e., a flat power density distribution across a broad band of frequencies), called transcranial random noise stimulation (tRNS) was first proposed by Terney and colleagues to desynchronize pathological cortical rhythms (Terney et al., 2008). Generally, tRNS uses a frequency range between 0.1 Hz and 640 Hz (full spectrum) or 100–640 Hz (high frequency stimulation). The lower boundary at 0.1 Hz was chosen to avoid DC effects, and the higher boundary was chosen according to the fast thalamic somatosensory evoked potential frequencies. The probability function of the stimulation follows a Gaussian curve with zero mean and a variance, where 99% of all generated current levels are within the target amplitude. It was shown that tRNS with frequencies between 100 and 640 Hz can increase the excitability in the motor cortex (Terney et al., 2008), measured by TMS.

It is still unclear if tRNS entrains resonance frequencies or works via stochastic mechanisms, via specific modulation of the excitatory-inhibitory balance in the brain or by an increase in synchronization by amplifying subthreshold activity (Fertonani and Miniussi, 2017). The animal studies using tRNS are missing. In humans, in a pilot study the Na+ channel blocker carbamazepine showed a tendency towards decreasing the size of MEP amplitude after the motor cortex stimulation (Chai et al., 2015). Compared to tACS, tRNS has a better blinding potential with less itching, tingling or burning sensations during stimulation (Ambrus et al.,
Furthermore, retinal phosphene perception in a wide frequency range (6–70 Hz) is a side effect of specifically tACS, but not of TRNS.

1) Recently new stimulation protocols were introduced with the aim to enhance the spatiotemporal precision and penetrability of TACS, reaching deeper brain areas (Grossman et al., 2017; Voroslakos et al., 2018). Using temporal interference stimulation (TI-TACS) the application of two electrically isolated currents at kHz frequencies (e.g., 2000 Hz and 2010 Hz), can temporally interfere deep in the brain to create an envelope amplitude that changes periodically at the slow difference frequency (Grossman et al., 2017). In the mouse, TI-tACS could recruit neural activity selectively in the hippocampus, without recruiting neurons of the overlying cortex. Whether this method can be used in human subjects, is still under investigations.

2) Other possibility is to apply time shifted multiple short pulses of currents, via different pairs of electrodes. This intersectional short pulse stimulation (ISPS) can be performed with pulses of 2.5 or 50 μs duration with 5 or 50 μs inter pulse interval (Voroslakos et al., 2018). By spatiotemporally rotating stimulation, deeper areas in rodents were reached. Application of ISPS in healthy human subjects modulated the amplitude of alpha activity in the visual cortex, recorded by EEG.

6.3. How can tACS be used in the clinic?

Although TACS has the potential to normalize maladaptive oscillatory activity in the human brain, the number of clinical studies applying this kind of approach is so far limited but has included tinnitus, depression and schizophrenia. One of the most successful approaches has been modulation of network oscillations in schizophrenia (Ahn et al., 2018). In a randomised double-blind, sham-controlled clinical trial, schizophrenia patients with auditory hallucinations received twice-daily 10 Hz-tACS for 5 days. After treatment, clinical improvement of auditory hallucinations correlated with enhancement of alpha oscillations.

TACS might be a treatment option for patients suffering from tremor in Parkinson disease (PD). Brittain and coworkers (Brittain et al., 2013) applied TACS over the motor cortex in patients diagnosed with tremor-dominant PD. The oscillatory activity responsible for tremor is thought to arise from a network of brain structures including cortex, basal ganglia and cerebellum (Helmich et al., 2011). TACS was most effective at the individual tremor frequency for inducing cortical phase cancellation, presumably due to suppression of the resting tremor amplitude. However, this study used an almost closed-loop stimulation setup in which the tremor frequency was measured online and motor cortex stimulation parameters (phase) were adjusted according to the measured activity, a setup that is difficult to realise. Nevertheless, individually adjusted closed-loop stimulation can considerably surpass the efficacy of open loop approaches. In a recent study (Guerra et al., 2021) PD patients were stimulated over the M1 using β- and γ-tACS and the kinematic features of repetitive finger tapping were analysed. The movement velocity significantly worsened during β-TACS and movement amplitude improved during γ-TACS. The effects of TACS were comparable between OFF and ON sessions and there was a positive correlation between the effect of γ-TACS on movement amplitude and motor symptoms severity.

Recent data support the feasibility of utilizing TACS to prevent cognitive decline in elderly. tACS applied to the left prefrontal cortex (PFC) and left temporal cortex at a theta-band frequency was shown to improve performance in working-memory tasks of elderly people (Reinhart and Nguyen, 2019).

In summary, compared to other transcranial stimulation methods (TMS or tDCS), the field of tACS is still in its infancy. Since the first systematic human research TACS study published (Antal et al., 2008), the method has been advanced in many ways; nevertheless, because of the complexity of the parameters, there are still concerns about the methodology and the efficacy of the stimulation. The following issues should be considered during planning a TACS experiment: 1) Which brain process shall be modulated and which oscillations are associated with this process? 2) What is the targeted outcome and which parameters of the brain oscillation (frequency, phase, coherence, amplitude) should be modulated to achieve the optimal effect? 3) Is it possible to simulate the current flow? (Modelling the intracranial current densities or referring to existing modelling results?) Can the desired effect be modelled in the neuronal network? 4) Which brain area is targeted, which electrode montage should be used? 5) Is it possible to demonstrate both behavioral and physiological changes during or after stimulation? 6) Does a plausible control condition (sham or an additional real stimulation condition) exist to demonstrate the specificity of the stimulation? (Control stimulation frequencies should be chosen outside of harmonics).

Development of new, hypothesis-driven approaches based on brain oscillations, physiological and/or behavioral measurements are expected to facilitate progress in the near future.

7. Reaction time

Reaction time (RT) is simply a measure of the amount of time that elapses between the onset of a particular stimulus and the start of an associated voluntary response. While this may seem like a relatively crude and simplistic metric of human performance, this belies its true complexity in terms of the processes and structures that contribute to the speed with which a person can react. Many subcomponents of the central and peripheral nervous systems must interact and pass their outputs to an effector which will often be constrained by its own physical properties or by the environment. Below we will discuss how different RT paradigms and tasks can be used to gauge particular neural processes, how some recent studies have used RT to provide new insights into the nature of several neurophysiological disorders, and finally how measurement of RT is accomplished and interpreted.

7.1. Processing stages involved in reaction time

From an information processing perspective, the steps required in a RT task involve detection and discrimination of a stimulus, decision making regarding the appropriate response, planning and organization of the response, and finally response initiation and execution (Welford, 1988). In the 19th century F.C. Donders was among the first to link RT to the notion of mental chronometry (1868/1969). He reasoned that all other things being equal, any differences in RT could be attributed to the durations of the mental processes required to execute a particular task. In his experiment he contrasted three types of tasks that each required a different number of processing steps: A) simple RT, where a single stimulus was associated with a single known response; B) choice RT, where the required response depended on the stimulus that was presented; and C) discrimination RT (or go/no-go) where the identity of the stimulus instructed whether or not a single response was to be initiated. He argued that task “A” required only detection of the stimulus and initiation of the response, while task “C” additionally required discrimination of the identity of the stimulus. Task “B” further required a choice of which response was to be made (see
Donders suggested that subtracting the RTs between tasks could provide a direct measure of the time required for mental processes such as decision making and discrimination (1868/1969). For example, Donders suggested that subtracting RT in task C (go/no-go) from that in task B (choice RT) provided a direct measure of the time required for the mental processes related to making a response choice. Take the example of an Olympic athlete ready to start the 100 m sprint. In order to gain an advantage, the sprinter must react to the sound of the starting gun as quickly as possible, but not before, which would result in a false start. Because the sprinter knows the action that must be executed, they only need to detect and identify the relevant stimulus before initiating the action. Donders argued that this “simple RT” situation provided a pure measure of stimulus detection and response initiation time (see also Haith et al., 2016). Of course, we are now aware that the processes and functions involved in these types of responses are not so clear-cut, but the basic premise that RT can be used to gauge or index the speed of neural processes is still used today. As such, there are many different RT paradigms that have been employed throughout the literature to gain insight into the integrity and duration of particular mental processes in both healthy and clinical populations.

7.1.1. Simple reaction time paradigms

In simple RT tasks, some of the processing steps (e.g., response selection, response programming) can be completed in advance of the go-signal because the required response is known in advance (see Fig. 14). Consequently, simple RT presumably requires the fewest processing steps after the go-signal (i.e., during the RT interval), including only stimulus detection and response initiation. Therefore simple RT has been used as a measure of general processing speed (Jensen, 2006; Schubert et al., 2017) and it has been shown to be strongly correlated with intelligence (Jensen, 2011; Rijjsdijk et al., 1998). Because simple RT does not involve some of the processing requirements of more complex RT tasks discussed below, simple RT paradigms can be specifically tailored to separately investigate signal detection and response initiation processes. For example, simply changing the modality of the imperative go-stimulus (e.g., visual versus acoustic) leads to RT differences (Woodworth, 1938) as the different stimuli are transduced and processed by different neural structures (Brebrner and Welford, 1980; Luce, 1986). Accordingly, simple RT can be used to gauge any deficiencies in the processing of a particular sensory modality as compared to normative values. Processes related to response preparation can also be indexed using simple RT. Indeed, shorter RT latencies have long been associated with heightened preparatory state (Carlsen et al., 2012; Näätänen, 1971), and shorter simple RTs are correlated with a larger motor related potential measured using electroencephalography (EEG) over pre-motor and motor cortical areas (Hillyard, 1969). In a similar manner, low task engagement and fatigue can negatively affect simple RT through direct or indirect impacts on stimulus identification and response initiation processes (for additional information see Jahanshahi, 2003). It is important to note that simple RT is also facilitated by factors such as increased stimulus intensity (Kohfeld, 1971; Woodworth, 1938) and concurrent presentation of redundant stimuli (Maslovat et al., 2018b), which tend to speed processing of stimulus detection and response initiation processing respectively. In summary, it can be seen that even simple RT can provide an easy, non-invasive method for assessment and measurement of the speed of a variety of mental processes.

7.1.2. Complex reaction time paradigms

More complex RT paradigms can be used to assess the duration and effectiveness of higher order mental processes. For example, RT tasks have been used to investigate executive function including (but not limited to) decision making, response selection and programming, inhibitory processing, and stimulus–response conflict. For example, Hick (1968) showed that RT increased by a constant amount each time the number of stimulus–response alternatives doubled, suggesting that the choice-related processing of ‘bits’ of information required to reduce uncertainty by half led to this linear increase in RT. Similarly, it has been consistently shown that motor responses that are more complex (e.g., more sub-components, longer duration) incur longer reaction times, presumably due to increased processing demands related to organizing and programming the response ahead of its execution (Carlton and Newell, 1987; Klapp and Maslovat, 2020). Interestingly, this effect is seen (albeit smaller) even when the response is fully cued and can be planned in advance, a result attributed to cerebellar implementation of response timing that must occur following the go-signal (Maslovat et al., 2018a). The speed and effectiveness of response inhibition has been investigated using a variety of tasks, several of which are commonly used in clinical settings and in neu-
ological assessments. In a stop-signal RT task a countermanding signal is occasionally presented at a particular delay following the go-signal (Logan et al., 1984), and participants are instructed to try to inhibit the response if a stop signal occurs. The probability of response inhibition, along with RT on unsuccessful stop trials, allows for calculation of the latency of the covert stopping process (Logan et al., 2014). Importantly, this stop-signal RT can be used as a metric of inhibitory control. Similarly, the Stroop task (Stroop, 1935), which requires participants to name the colour of presented words, is often used in screening for brain damage and clinical assessment of both response inhibition and selective attention (for a meta-analysis see Faria et al., 2015; Homack and Riccio, 2004). A conflict arises when a colour word is presented in a colour different to that which it is naming (e.g., “RED” presented in blue text, where the participant must respond by verbalizing “blue”), resulting in delayed and incorrect responses. It is postulated that the word name automatically activates the word representation of that colour, which creates interference that must be consciously inhibited. While the tasks described above represent many of the commonly used RT paradigms, other RT tests such as the Simon task (used to measure inhibition of automatic responses at the response selection stage; see Lu and Proctor, 1995), dual task paradigms (used to assess attentional load; Giroud et al., 1984), vigilance tasks (used to assess sustained attention; see Dorian et al., 2005), and serial RT tasks (used to assess attention and implicit learning; Schwarz and Schumacher, 2012) are also often employed for assessing executive function. Finally, it should also be noted that within RT tasks, other measures such as response accuracy and error type can also provide valuable information regarding impacted processes. For example it was noted above that go/no-go RT tasks can be used as an index of stimulus discrimination time; however, the rate of commission errors (go on no-go stimulus) or omission errors (no-go on go-stimulus) can also provide insight into response inhibition and vigilance processing (Staub et al., 2014).

### 7.2. Reaction time and clinical neurophysiology

Based on the preceding sections it is clear that RT can provide insight into the functional integrity of various neural processes. Also, because it is a reasonably easy to obtain metric, the task variations can be tailored to identify and assess particular functional deficiencies in different clinical populations. While this is not an exhaustive list of uses of RT in clinical assessments, it is intended to highlight some of the ways in which RT can be used to identify the neural processes that are more impacted by a particular neurological condition, or more improved by a particular intervention. For example, RT has been used in order to identify which processes lead response slowness (bradykinesia) in Parkinson’s disease (PD). The general consensus of many different studies incorporating RT suggests that response initiation processes are most impacted in PD (Carlsen et al., 2013; Jahanshahi, 2003). Similarly, performance on stop-signal tasks is often impaired in individuals diagnosed on the impulsive-compulsive spectrum with disorders such as obsessive–compulsive disorder (OCD) and attention deficit hyperactivity disorder (ADHD), suggesting impairments specific to inhibitory function that also generalize to motor reactions (van Velzen et al., 2014). In patients with Multiple Sclerosis (MS), RT deficits specific to auditory and visual stimulus processing have been shown (Demaree et al., 1999) as have deficits in cognitive decision making processes at various stages of disease (Schulz et al., 2006). RT has also been used to assess severity of neural trauma and to predict time to recovery. For example, slowed RT is a particularly sensitive measure related to cognitive impairment following concussion (Collie et al., 2006; Warden et al., 2001). Moreover, better performance in various RT tasks has been shown to correlate with shorter time to recovery following concussion (Lau et al., 2009). RT has also been used as a predictor of recovery following stroke. For example, one study assessing simple and choice RT following acute stroke showed that faster performance on a simple RT task administered within two weeks of the stroke was related to the higher scores on the Montreal Cognitive Assessment (MoCA) and attentional function administered three months post stroke (Cumming et al., 2012). Finally, RT has been extensively used to assess the efficacy and impact of various motor and cognitive interventions on different neurological disorders, as well as on aging. For example, it has been shown that following four weeks of mirror therapy (five days per week) in patients with chronic stroke, RT and Fugl-Meyer scores showed greater improvement compared to conventional upper limb therapy (Wu et al., 2013). A recent systematic review suggested that the use of transcranial direct current stimulation (tDCS) as an adjuvant therapy in individuals with PD appeared to improve upper limb motor function as well as RT, but these effects were dependent on stimulation site and task (Simpson and Mak, 2020). Finally, a recent meta-analysis confirmed that step training focussing on executing rapid and accurate stepping motions, led to improvements not only in measures of gait and balance in older people, but also in measures of RT (Okubo et al., 2017).

### 7.3. Quantification and measurement of reaction time

Measurement of RT should be relatively straightforward considering it simply involves recording the amount of time that elapses between the presentation of an imperative stimulus and the initiation of the required response. However, it is important to reflect upon several factors related to recording and assessing RT that can impact the outcome. First, once a RT testing paradigm is decided upon, the characteristics of the RT stimulus must be considered. As noted above, intensity and modality of the stimulus plays a role, but other stimulus-related factors can also impact RT such as the timing of stimulus presentation. Often a foreperiod with a randomly variable duration is used between presentation of a warning (get-ready) signal and the imperative go-stimulus in order to discourage anticipation (“catch” trials, where no imperative stimulus occurs, are also used). In these cases RT tends to follow a “U” shaped curve whereby shortest (i.e., fastest) RT tends to occur in the middle of the range of possible foreperiods, with longer RTs at each extreme (Drazin, 1961; Niemi and Naätänen, 1981). The identity of the stimulus can also play a strong role in the outcome observed. For example high body mass index (BMI) individuals are faster to respond to images of food stimuli (as compared to non-edible objects), and when used in a go/no-go paradigm, more commission errors were made when the food was the “no-go” stimulus (Meule et al., 2014). Oftentimes many of these factors do not need extensive consideration in clinical settings because one of most widely employed methods for measuring RT involves a keypress or mouse click in standardized computer-based tasks (e.g. CANTAB, see Lenehan et al., 2016), although object “drop” tests have also been used (Eckner et al., 2012). Newer online virtualized neurological batteries have also begun to be used which also include RT assessment (de Leeuw and Motz, 2016; Domen et al., 2019), with some normative results being crowdsourced online (Bazilinskyy and Winter, 2018). Yet, even in computer-based assessment, differences in computer hardware characteristics can impact measurement of RT, with tablet-based devices showing variable timing delays of 20–100 ms (Schatz et al., 2015). The actual “response” being measured can also have a material impact on RT (Brener and Smeets, 2019). Detection of movement can be measured using a key-press (or release), goniometers, force transducers, accelerometers, optical / magnetic motion capture devices, onset of electromyo-
sion is that reaction time is often conflated with “response time,” which not only includes the time to initiate the response, but also the time to response completion. Once the movement command has been released at the cortical level, nerve conduction time, muscle activation time, and electromechanical delays (e.g., time to overcome elastic and inertial properties of a particular limb) can all contribute to RT delays and variability (e.g., Winter and Brookes, 1991). Indeed, even in a task requiring a simple finger lift, it has been shown that inclusion of additional “movement time” to lift the finger off a button (as compared to measuring the onset of change in force) could impact both the estimated RT and the conclusions reached about differences between conditions (Brenner and Smeets, 2019). This differential would only be heightened with more complex tasks such as sit-to-stand, or other multiple-component actions. Thus, care should be taken so that RT measurements reflect the earliest possible detection of movement onset to best reflect the duration of central processes. In order to obtain an even more accurate measurement of central RT, researchers have long employed electrophysiological methods to further “fractionate” RT (Botwinick and Thompson, 1966) in order to identify markers that allow for the separation of the durations of different processes that occur during response onset (see Fig. 15). For example, RT can be split into premotor RT - which entails the amount of time between the stimulus and onset of electromyographic (EMG) activity in the target effector (marker 3, Fig. 15), and motor time - which characterizes the time from EMG onset to overt movement detection. As such, premotor RT is considered to be more representative of the duration of central processes, with peripheral effects removed. RT can further be fractionated at the cortical level by using transcranial magnetic stimulation (TMS) to detect the rise in corticospinal excitability preceding response onsets. Finally, EEG measures such as the lateralized readiness potential (Leuthold et al., 1996), and event related desynchronization in the beta band (averaged over many trials) can provide some insight into the end of response selection processes in cortex (Leocani et al., 2001) (see Fig. 15). In sum, the methods chosen to present stimuli and to measure and assess the onset of the response can have meaningful impacts on the RT measured as well as the interpretations made based on those measurements, thus careful consideration should be given to all aspects of experimental design when employing RT as a dependent measure.

8. BP (Bereitschaftspotential), CNV (Contingent Negative Variation)

This section will deal with the movement-related potentials (MRPs). MRPs are Event Related Potentials (ERPs) associated with the processes of voluntary movement preparation, initiation and execution in different paradigms. Several types of MRPs have been described and they are all slow negative-going potentials. The Bereitschaftspotential (BP) is discernible 1 to 2 seconds prior to the execution of a voluntary movement. In the 1980s, a related discovery was the lateralized readiness potential (LRP) as a measure of pre-movement EEG asymmetry from which the presence of preferential preparatory activity can be inferred. The Contingent Negative Variation (CNV) is recorded when a warning stimulus (WS or S1) precedes an imperative or response stimulus (RS or S2) signalling the requirement of a quick response. A related ERP is the Stimulus-Preceding Negativity (SPN) which becomes evident when a person is anticipating a forthcoming salient stimulus bearing significant ‘knowledge of results’ information. The SPN which is observed without a motor response, will be briefly discussed in relation to the CNV, as strictly speaking, it is not an MRP (Brunia et al., 2012). Surface negative potentials such as the BP and the CNV are considered to represent neuronal activation (Skinner and Yingling, 1977) and the BP to specifically represent excitatory post-synaptic potentials (Caspers and Speckmann, 1974).
The BP (Fig. 16) was first described by Kornhuber and Deecke (1964) in their seminal paper on the readiness potential. Kornhuber and Deecke (1965) stated that “Voluntary hand or foot movements are preceded by a slowly increasing surface negative cortical potential of 10–15 μV, called readiness potential. This potential is maximal over the contralateral precentral region but shows bilateral spread and is larger over the frontal than over the occipital areas. The readiness potential increases with intentional engagement and is reduced by mental indifference of the subject.” In a typical case the subject is asked to make a series of self-paced, random movements (e.g. approximately one response every 5 s) with either the upper (e.g. button press), or the lower limb (e.g. pedal press). No stimuli are needed as the impending voluntary movement is a sufficient condition for the emergence of the BP and the BP is time-locked to movement onset (Brunia et al., 2012; Eimer and Coles, 2002; Shibasaki and Hallett, 2006). More recently, the BP has also been recorded prior to naturalistic movements such as a 192-meter bungee jump (Nann et al., 2019).

Typically, the BP is recorded from surface, scalp electrodes by the use of the EEG, although it can also be recorded from single neurons in the human medial frontal cortex (Fried et al., 2011), or from deep brain structures, the basal ganglia (caudate, putamen and pallidum) and the thalamus (Eimer and Coles, 2002; Rektor, 2002). Using MEG, a ‘Bereitschaftsfeld’ equivalent of the BP has also been recorded (Deecke and Kornhuber, 2002). With the improved temporal resolution of event-related fMRI, Cunnington et al. (1999) have recorded a ‘Bereitschafts-BOLD’ response, the BP equivalent in the hemodynamic response. As a historical note, when Kornhuber and Deecke (1964) made the first BP recordings they had to do off-line back-averaging of the EEG segment prior to the EMG by playing the tape backwards. Modern technology allows for easy recording of the EEG and easy averaging procedure of, typically at least 20 self-paced movements.

A few distinct components can be discerned during the course of the BP (Fig. 16). The first part of the BP, starting 2 to 1 s before a movement, is the so-called ‘early BP’, and has a more diffuse, yet mainly midline distribution over the cortex. The early BP is thought to reflect more general preparation for the forthcoming movement (Eimer and Coles, 2002; Shibasaki and Hallett, 2006) and its generation has been linked to the pre-supplementary motor area (pre-SMA), supplementary motor area (SMA) and the lateral premotor cortex bilaterally corresponding to the Brodmann area 6 (Brunia et al., 2012; Shibasaki and Hallett, 2006). There is evidence from current source density analysis (Ball et al., 1999; Cui et al., 1996), concurrent MEG and PET (Pedersen et al., 1998), and event-related fMRI studies (Cunnington et al., 1999; Cunnington et al., 2002) that during the BP recorded prior to self-paced movements, pre-SMA and SMA activation precede motor cortex activation. The early BP is followed by the ‘late BP’ (Shibasaki and Hallett, 2006), starting 400–500 ms before the movement, characterized by a sudden shift of the gradient of the negativity at the central electrodes contralateral to side of movement. This late BP has been related to activation of the primary motor cortex (Brunia et al., 2012). There is indeed evidence that these two BP components are functionally related to different brain areas. In the literature different terminology has been used to refer to these earlier and later phases of the BP (Eimer and Coles, 2002). While ‘early BP’ has been variably referred to as simply ‘BP’, ‘BP1’ (Deecke et al., 1969), or negative slope 1 (NS1); ‘late BP’ has been referred to as BP2 (Deecke et al., 1969), negative slope (NS) (Shibasaki and Hallett, 2006), negative slope 2 (NS2). There are also two components following late BP – the premotor positivity (PMP), seen 50 ms before the movement (also referred to as P-50), and the motor potential (MP) occurring 10 ms before the movement onset (hence, the designation N-10), as well as the post-motor potentials (Shibasaki and Hallett, 2006). PMP is predominant over the hemisphere ipsilateral to the moving hand, and it is not seen in simultaneous bilateral hand movements. Therefore, it might represent a suppression of the movement of the opposite hand in intended unilateral hand movements such as during physiological mirror movements (Shibasaki and Hallett, 2006). Another explanation is that it does not have any specific function, and it is just a trough between two negative peaks without any physiological significance. MP is localized to a small area of the contralateral central scalp corresponding to the movement site and occurs immediately before the movement onset. This component most likely represents the activity of the pyramidal tract neurons in the primary motor cortex (M1). There are also a few post-movement potentials (N + 50, P + 90, N + 160 and P + 300) which will not be further discussed as our focus is on the premovement ERP components.
Regarding the subcortical sources of BP, it is likely that the activation of cortical regions associated with BP is triggered by subcortical areas such as the basal ganglia, thalamus, and dentate nucleus of the cerebellum (Brunia et al., 2012; Ikeda and Shibasaki, 2002). All of these structures play an important role in networks responsible for the planning and execution of voluntary and involuntary movements, but also in response selection and inhibition of competing motor programs (Brunia et al., 2012; Jahanshahi et al., 2015). BP recordings have been obtained from numerous subcortical regions during surgery for refractory epilepsy (Rektor, 2002). During these operations, depth electrodes are used to determine the exact location of the epileptic focus. In addition to cortical sources (pre-SMA, SMA, primary motor cortex, cingulate gyrus), this approach also revealed that a BP can be recorded from the pallidum, putamen, and head of the caudate nucleus (Rektor, 2002). A BP was also recorded from the thalamus (Paradiso et al., 2004) in six patients with essential tremor and one patient with myoclonus dystonia who underwent deep brain stimulation. The authors were able to record bipolar pre-movement potentials with the same time delay to movement onset as those recorded by the cortex BP (Brunia et al., 2012; Paradiso et al., 2004). In four patients, the maximum amplitude was recorded from the ventral lateral nucleus of the thalamus. Phase reversal occurred in four of the recordings, indicating a local source of generation rather than simple volume conduction from other (e.g., cortical) sources. In patients with cerebellar lesions, particularly those involving the dentate nucleus, BP could not be recorded (Shibasaki et al., 1978). Moreover, the dentato-thalamocortical pathway seems to be crucial for the generation of BP (Ikeda and Shibasaki, 2002; Shibasaki et al., 1978). In summary, all of these subcortical and cortical structures, from which a BP can be recorded, could be part of the movement-preparatory networks that are important for the generation of the cortical BP.

A further related discovery in the 1980s was of the lateralized readiness potential (LRP) (Fig. 17); the discovery related to the observation that in a choice reaction time task requiring left or right responses to an imperative stimulus, the time point when brain activity became asymmetrical related to the time when participants were provided with information about whether to respond with their right or left hand; leading to the conclusion that the asymmetry in the EEG record was an index of preparation to execute a specific hand movement (Kutas and Donchin, 1980). Following on from this observation, the groups in Illinois (Coles and Gratton, 1986) and in Groningen (Smid et al., 1987) developed procedures for deriving the LRP, as a measure of pre-movement EEG asymmetry from which the presence of preferential preparatory activity could be inferred. The LRP is calculated by subtracting the ipsilateral from the contralateral EEG activity prior to movement (Eimer and Coles, 2002) (see legend of Fig. 17). With the LRP it is possible to more precisely examine the cortical activity related to preparation and initiation of a contralateral limb movement. The LRP is a measure of the lateralised asymmetric section of the BP. The LRP has become an important measure of covert cognitive processing and has been employed to investigate a host of...
effects and processes including stimulus–response compatibility and cue validity effects, the Simon effect, response inhibition, subliminal perception, implicit learning, serial versus parallel models of information processing (Eimer and Coles, 2002). More recently, using a motion discrimination task, it has been shown that the LRP is associated with and reflects the crossing of a decision threshold (van Vugt et al., 2014).

Since the BP precedes movement by 1 to 2 s, it was portrayed as a readiness potential, that is an index of motor preparation (Kornhuber and Deecke, 1965). The amplitude and latency of the BP recorded are influenced by various factors (Shibasaki and Hallett, 2006). Not only simple movement parameters such as force (Kutas and Donchin, 1980) and rate (MacKinnon et al., 1996) impact the amplitude and latency of the BP; but also higher order motor processes such as movement complexity (Benecke et al., 1985; Simonetta et al., 1991), effort (Cunnington et al., 1999; Deecke and Kornhuber, 2002; Wessel et al., 1994) and mode of movement selection (Dirnberger et al., 1998; Jahanshahi et al., 1995; Praamstra et al., 1996; Touge et al., 1995) are important.

The reliability of action consequences, whether consistent or inconsistent, has also been demonstrated to influence the amplitude of both the early and late BP (Wen et al., 2018). In addition to being defined as a readiness potential, based on the types of tasks and different experimental manipulations, the BP has been considered to reflect a host of other processes. For example, it has been suggested that the BP reflects the unconscious preparation and initiation of action (Libet et al., 1983), an index of resource mobilisation (Eimer and Coles, 2002), timing of movements (Deecke et al., 1985) or evidence accumulation to an internal threshold for action (Schurger et al., 2012; Travers et al., 2020). If the argument that the late CNV and BP are identical is accepted (Brunia et al., 2012), then the set of cognitive, motivational and motor processes such as anticipation and expectancy, attention, preparatory set, time estimation, information processing, and motivation which are considered to be reflected by the CNV could also be pertinent for the BP (Brunia et al., 2012; Eimer and Coles, 2002). In fact, some have suggested that non-motor decision-related or anticipatory processes are reflected by the BP (Alexander et al., 2016).

There is also evidence that distinct variables modulate the amplitude of the early and late BP (for a review see Jahanshahi and Hallett (2002b) and Shibasaki and Hallett, 2006). While precision, discreteness and complexity of movement are associated with larger late BP, they have no effect on the early BP (Shibasaki and Hallett, 2006). By contrast, higher intention to move, learning, mode of movement selection are associated with greater early BP, but have no effect on the late BP (Shibasaki and Hallett, 2006). The degree of regularity/irregularity of movement, which influences the extent of motor preparation also affects the early BP (Jahanshahi et al., 1995). In healthy people and as well as people with Parkinson’s disease dopaminergic medication increases the amplitude of the early BP, it has no effect on the late BP (Dick et al., 1987). Neurological disorders such as Parkinson’s disease and cerebellar disease also differentially impaire the BP components, suggesting that the striatal-frontal and cerebello-frontal circuits primarily contribute to specific BP components. Off medication, the amplitude of the early but not the late BP is reduced in Parkinson’s disease (Dick et al., 1989; Jahanshahi et al., 1995; Shibasaki et al., 1978), for a review see Georgiev et al. (2016). By contrast, in people with cerebellar degeneration or atrophy, the late BP is absent or reduced (Ikedo et al., 1994; Shibasaki et al., 1978; Tarikka et al., 1993; Wessel et al., 1994).

In general, compared to the BP amplitude, latency has been found less informative. However, the preparatory state influences BP latency such that movements with some general preplanning or preparation to act in near future have earlier onset (Shibasaki and Hallett, 2006). The speed of movement also affects the latency of BP in a proportionate way – the faster the speed of movement, the later (closer to the movement onset) the BP begins. This seems to be counterintuitive, but one should remember that the BP starts before the movement, which means that higher speed shortens the duration of the BP by moving it closer to the time of execution (Shibasaki and Hallett, 2006).

Early studies by Libet measuring the BP with the ‘clock’ paradigm (Libet, 1992; Libet et al., 1983) have suggested that conscious awareness of an intention to act occurs some time, on average 350–850 ms after the onset of the BP. The precise latency depends on whether the movement was perceived as pre-planned or spontaneous, as well as on the method used to estimate onset of awareness. These findings implied that the decision/intention to act and the preparation for and the initiation of movement as indexed by the BP occur unconsciously. The inference has been that in order for the movement to be consciously perceived the BP negativity needs to reach a certain amplitude at a certain time point in the course of movement preparation. This is supported by the fact that automatic and involuntary movements are not preceded by a BP. However, Schlegel et al. (Schlegel et al., 2013) failed to find any correlation between the onset of the BP and the time of awareness to move, suggesting that the processes involved in production of the BP and conscious awareness may be separate. Some confirmation for a delay before decisions reach awareness was provided by...
an fMRI study requiring free decisions about right- or left-hand movements, with the time of intention to move being quantified using presentation of a stream of letters (Soon et al., 2008). The results showed that a decision can be encoded in prefrontal and parietal activity up to 10 s before entering awareness. It was proposed that the delay may reflect the preparation of high-level control networks before the decision enters awareness (Soon et al., 2008).

The Libet findings had a profound implication for our understanding of “free will” in the last three decades, suggesting that “free will” might actually be restricted to selecting or vetoing action prior to completion. The findings were also interpreted in a way that many of the so-called “voluntary” actions may be unconsciously initiated which might be a means of liberating conscious processes for more important control mechanisms. The results also questioned the causal relation of our decisions/actions to our intentions. If our intentions drive our actions, then what would be the role of consciousness and are there any other ways of action generation in addition to the intention-action causal route? The Libet experiment (Libet et al., 1983) was actually based on the difference of subjectively perceived time of the urge to move (W) and the time of actual movement (M). However, recently Dominik et al. (2017) have replicated Libet’s experiment, but the participants were naive regarding the nature of the experiment and performed the W and M judgments in two different orders. When W judgments were made before M, there was no actual difference in the perceived time between W and M across participants, suggesting that in Libet’s experiment ‘W’ was not based on introspection but partially inferred from previous experience with M, which casts doubt on the validity of the intention reports (Dominik et al., 2017). From the results of a further study replicating, Libet’s experiment, Schurger et al. (2012) have proposed that the BP can be explained by a leaky stochastic accumulator model and that the BP is a by-product of a drift diffusion process such that movement onset is determined by crossing a threshold set on spontaneous fluctuations of neural activity. Based on their EEG findings the BP was given a true mechanistic role of serving as a neural accumulator and a threshold in terms of ongoing spontaneous fluctuations in neural activity. Moreover, it might be that the activity designated as BP is purely stochastic and it could be found even without preceding movement (Schurger et al., 2012) although this has recently been questioned (Travers et al., 2020). Recent data suggests that people can veto and stop a movement even after the onset of the BP (Schultze-Kraft et al., 2016). Regarding clinical applications, other than its recording in movement disorders such as Parkinson’s disease, dystonia and cerebellar disease, the BP back averaging has mostly been used to discern ‘functional’ or ‘psychogenic’ from other movement disorders in individual patients. Myoclonic jerks are involuntary movements. Myoclonus can be a manifestation of different neurological disorders and a wide palette of medical conditions with a clear organic basis, but it can also be ‘functional’ in origin (Terada et al., 1995). If a premovement negativity consistent with a BP is present on back-averaging in a patient with myoclonic jerks, then these abnormal movements are either ‘psychogenic’ or tics (Duggal and Nizamie, 2002; Karp et al., 1996; Obeso et al., 1981; Vial et al., 2019). In tic disorders, the BP is reported to be shorter than normal (Duggal and Nizamie, 2002; Karp et al., 1996).

8.2. The contingent negative variation

The CNV is a slow negative brain potential that develops between two consecutive stimuli, WS (S1) and RS (S2), where S1 is a warning stimulus anticipating the imperative or response stimulus S2 that signals the need for initiation and execution of a motor response. Therefore, the CNV represents the neural activity necessary for sensorimotor integration or association and is related to preparation and execution of externally-paced, voluntary movements (Brunia et al., 2012). Unlike the BP that reflects the preparation of voluntary self-paced movements, the CNV reflects preparation for externally-triggered or signalled movements (Brunia, 2002). To an extent, both, the BP and CNV represent processes of anticipation and preparation, the major difference being that for the CNV the anticipation is conditioned by an external signal. Theoretically speaking, even voluntary movements are triggered by some kind of event or activity, i.e. a form of internal ‘stimulus’ or decision, albeit not external. Nevertheless, anticipatory behaviour as reflected by the CNV is not restricted to the motor system, it also involves attention to the external stimuli, which means that the CNV heavily depends on attentional, that is cognitive processes.

The CNV was first described in the sixties (Walter et al., 1964) (Fig. 18). In this study, the WS was a simple click and the RS was a series of flashes with an interstimulus interval of 1 s. They noticed a rise in negativity after the WS that lasted until the subject responds and presses the button after the presentation of RS. As it was described in the original paper: “This pattern is maintained indefinitely as long as the subject is attentive and presses the button promptly” (Walter et al., 1964). This statement was an early indication that the preparation of movement is not the only factor that determines the CNV, but also attention. Jarvilehto and Frustorfer (1970) were the first to suggest that there might be distinctive components related to the CNV – a frontal dominant, early, non-motor component, related to the properties of the WS, and a central, late, motor component, related to response preparation for the RS. The late component is similar to the BP. These two components are much easier to discern when an interstimulus interval longer than the originally used 1 s is used to elicit the CNV.

The CNV is easily elicited by a forewarned simple reaction time (SKT) task (Brunia, 2002). However, other tasks, such as a choice reaction time (CRT) task can also be used to elicit the CNV, introducing event uncertainty in the generation of the potential resulting in a smaller late-wave amplitude compared to the SRT. Similarly, if changeable foreperiods/interstimulus intervals are presented (time uncertainty), the late wave amplitude is also smaller. By contrast, if muscular effort is needed for a response to the RS, the amplitude of the late CNV wave increases. Speed instructions also affect the amplitude of the CNV, faster responses being associated with higher amplitudes. If a task is more perceptual than motor, or even only perceptual, then the CNV amplitude is much smaller (Brunia, 2002). In these cases, the scalp distribution also changes and shifts from frontal to parietal predominance.

Evidence indicates that premotor and prefrontal cortices, including the SMA, as well as the basal ganglia are important in generation of this cortical activity. The early CNV component is more frontally distributed and involves the prefrontal cortex, SMA and cingulate cortex and is linked to the arousal and attention associated with the WS. The late CNV has a more central distribution and CNV-like activity can also be recorded from the putamen, implying a crucial importance of the basal ganglia-thalamocortical circuits in CNV generation as well (Brunia et al., 2012). Pharmacologically, the most explicit model for the CNV states that its amplitude is determined by the activity of cholinergic neurons, which are in turn influenced by other neurotransmitters – dopamine, noradrenaline and gamma-aminobutyric acid (Brunia et al., 2012).

In a clinical context, patients with PD have a smaller CNV amplitude compared to health controls (Pulvermüller et al., 1996). The same holds true for patient with cerebellar atrophy (Verliger et al., 1999), although a previous study (Ikeda et al., 1994) described a dissociation between the BP and CNV in a patient with...
a small infarct in the mesial tegmentum involving the decussation of the superior cerebellar peduncle. In this study, the BP could not be recorded, but the CNV was normally elicited, probably due to the fact that the mechanism of cerebellar dysfunction was related to a lesion located in the mesencephalon, rather than a lesion of the cerebellum itself. However, this study is important in that it suggested different mechanisms of generation of the two potentials, with greater involvement of attentional processes in the CNV.

Smaller amplitudes of CNV have also been described in psychiatric disorders, such as schizophrenia (Klein et al., 1996; Wagner et al., 1996) and phobia (Regan and Howard, 1995; van den Bosch, 1984). A recent study has suggested that the CNV can be useful as a biomarker of attention in functional movement disorders (Teodoro et al., 2020). In this study, clinical improvement of functional movement disorders after physiotherapy was associated with faster reaction times and normalization of CNV, which was absent at baseline, suggesting that the CNV may be a useful neurophysiological biomarker related to abnormal attention in functional movement disorders.

A disadvantage in a warned SRT/CRT is that the attention for the upcoming stimulus and preparation for the movement take place simultaneously, such that the late CNV concurrently reflects both processes, which are thus confounded. Another paradigm, the time estimation task, is used to separate in time motor preparation from anticipatory attention. In this task, the participants have to press a button 2 s after an instruction stimulus and 2 s after the button press, feedback is presented about the correctness of the response, using knowledge of results (KR) stimulus. The responses (button press being too early or too late) are recorded and time-locked and then analysed. The KR stimulus elicits a Stimulus Preceding Negativity (SPN) (Brunia, 1988; Brunia, 2002), Fig. 19. The distribution of this potential is more frontal, similar to the early, WS-related CNV component. The SPN was introduced to differentiate between the CNV and the BP as a true non-motor CNV. In a very broad sense, the SPN can also be viewed as an index of attention. However, the interpretation of the functional significance of this potential depends on the type of stimulus used. It could therefore represent a basic anticipation of the stimulus, anticipation of the information content of the stimulus and emotional anticipation. As such, this potential is a non-motor anticipatory slow wave that is especially large preceding affective-motivational stimuli such as KR.

In conclusion, all the potentials reviewed represent anticipatory brain activity. While the BP and LRP are mostly MRPs, the CNV late wave is both a movement and perception related potential. SPN is a non-motor ERP component that is especially large preceding affective-motivational stimuli providing KR.
Fig. 20. Schematic overview of different types of motor learning. A: Motor adaptation. Starting with a well-established skill (pre), the learner is confronted with a modified environment that induces abrupt mismatch between the actual and the necessary execution (per). This leads immediately to high errors that decline relatively fast in a monotonic fashion back to baseline. When the environmental perturbation is removed, the learner continues with the adapted behavior that leads to errors in the opposite direction (after-effects, post). These errors decline rapidly back to baseline. The overall gain in skill is zero. B: Assisted adaptation. Starting with impaired performance (pre), an assistive device such as an orthosis or a cane, can improve behavior almost immediately due to its mechanical support (per). However, removing the device leads almost no after-effects as the learner has not adapted its unassisted behavior. C: Skill acquisition and retention. Starting with a lack of proficiency (pre), the learner gradually acquires skill in a successive practice sessions performance typically starts with an initial fast change followed by a slower asymptotic approach to expert level. In other scenarios, performance can improve discontinuously, for example when sudden insight reveals a better strategy to accomplish the task goal. Over successive practice sessions performance typically starts with some decrements compared to the previous performance, but then tends to improve faster reaching further improvements in each session. Practice also tends to be followed by off-line learning or consolidation, stabilizing a coordinative pattern after some interval of rest.

9. Motor learning: Adaptation and skill acquisition

9.1. Definition and scope

Motor learning is an extensive concept with many manifestations that are central to numerous disciplines. Motor neuroscience explores the fundamental mechanisms of neuroplasticity, developmental science examines the evolution of sensorimotor skills to assess, predict and treat potential neurological problems, and rehabilitation science aims to guide the re-learning of motor behaviors following neurological injury. As a first pass, master a motor skill is typified by decreasing variability and/or increasing accuracy and speed (Guthrie, 1935; Willingham, 1998). More rapid execution is often coupled with diminishing accuracy, acknowledged in the ubiquitous speed-accuracy trade-off (Fitts, 1954). This trade-off also shifts with practice as skilled individuals can become less variable, while keeping the same tempo, or they can move faster without increasing variability (Shmuelof et al., 2012).

These visible improvements in behavior can occur at different time scales: Proficient performance can be achieved right on the first trial or, more commonly, it requires exploration followed by a long period of honing the skill. An example for fast adaptation is when lifting a cup of coffee to drink, the changing weight of the content requires almost instantaneous adaptations over repeated lifts. In contrast, learning a difficult skill can take weeks or years, if not a lifetime, as easily appreciated in ballet dancers or professional pianists. Improvements tend to proceed gradually, typically starting with an initial fast change followed by a slower asymptotic approach to expert level. In other scenarios, performance can improve discontinuously, for example when sudden insight reveals a better strategy to accomplish the task goal. Over successive practice sessions performance typically starts with some decrements compared to the previous performance, but then tends to improve faster reaching further improvements in each session. Practice also tends to be followed by off-line learning or consolidation, stabilizing a coordinative pattern after some interval of rest.

Common to all manifestations of motor learning is the requirement that practice should lead to changes that persist beyond the practice session. Hence, evaluation of learning requires retention tests, typically after one or more days of no practice. True learning also implies transfer and generalization of the acquired skill to related tasks. Lasting performance changes reflect neuroplastic changes in the central nervous system.

9.2. Behavioral characteristics of motor learning

9.2.1. Motor learning, adaptation and skill learning

Motor learning is a multi-faceted phenomenon and an umbrella term for different learning scenarios with different underlying neural mechanisms (Fig. 20). One important distinction is between adaptation of well-established behaviors to altered environmental conditions, such as walking on different terrain, and the acquisition of a novel skill, such as learning to dance salsa. Adaptation epitomizes an ability that is ubiquitous and essential to all daily activities. It has received much attention over recent decades in experimental paradigms such as prism, saccadic, visuomotor, or force-field adaptations. For example, reaching to a target when exposed to a modified visuomotor mapping or to a force field acting on the hand requires recalibration of the previously well-known reaching movement (Redding et al., 2005; Shadmehr and Mussa-Ivaldi, 1994). Functional adaptations happen fast and typically errors decrease with a monotonic exponential time course. Fig. 20A illustrates the typical time course of adaptation: a well-established behavior (with zero error, ‘pre’) is confronted with a new task demand that abruptly induces an error (‘per’). However, the error quickly declines to re-establish baseline performance. Removing the perturbation leaves performance with errors in the opposite direction, known as after-effects, that again quickly vanish restoring baseline behavior (‘post’). Note that the error metric can take on many forms, determined by the skill. A different type of adaptation occurs when humans use assistive devices, such as wearing an orthosis or using a cane for the support of balance (Huber et al., 2019): the initial affected behavior is corrected...
almost instantly as the device imposes a mechanical change to the motor system (Fig. 20B). However, without this device, there is no after-effect or any lasting improvements — their effect on re-learning a behavior with neural effects is negligible. While such assistive devices are helpful, they should not be confused with therapeutic strategies that aim to re-establish the original skill.

These fast changes contrast with the process of acquiring a novel motor skill that was not in the person's natural repertoire (Fig. 20C); the learner progresses from the initially naïve state (high error) to a high degree of proficiency (zero error) in a typically monotonic fashion. These performance improvements are far slower and can take weeks or years. Importantly, such new coordinative skills do not return to the initial state, but persist, i.e., exhibit long-term retention (Park et al., 2013). One never forgets how to ride a bicycle.

9.2.2. Enhancing motor learning

Learning can be facilitated or accelerated by suitable practice conditions and training schedules to make the error decline faster and further enhance the increase in skill level (Fig. 20D). Evidently, this is the core of any coaching and rehabilitative therapy. Practice sessions that have proven most helpful include variations of the focal task that are practiced in a distributed manner (variable and distributed practice) (Schmidt et al., 2018). Massed practice and the attempt to maximize repetitions of the same movement, i.e., rote-learning, exhibits diminished retention and little generalization. Another proven means of enhancing learning is providing explicit quantitative feedback about performance parameters or about its final outcome (Salmoni et al., 1984). Giving reward or positive reinforcement has informational and motivational effects that has proven helpful, while negative reinforcement or punishment has less or even negative effects (Nikooyan and Ahmed, 2015; Wickens et al., 2003). The learner can also pick up a skill from observation alone, without being assisted by verbal explanation or by haptic guidance (Bandura, 2008; Burke et al., 2010). An important consideration when providing haptic guidance is that the learner needs to stay involved, i.e., produce a voluntary drive to activate muscles, as sheer passive guidance has minimal effect (Lotze et al., 2003). This insight has led to the ‘asist-as-needed’ principle in robotic therapy, where the robotic device assists only when the patient is unable to move their limb themselves (Marchal-Crespo and Reinkensmeyer, 2009; Pehlivan et al., 2016).

9.3. Computational approach to motor learning

From a computational perspective learning can be viewed as a directed process of optimization based on internal models navigating in a high-dimensional solution space or proceed in a model-free way, by trial and error. With practice humans develop an internal model of the body and the environment. These internal models have been considered as pairs of inverse and forward models required for the predictive (feedforward) control of actions (Wolpert and Kawato, 1998; Wolpert et al., 1998). Fig. 21 illustrates the basic flow of information. Upon initiation of a movement, the inverse model comprises neural processes that map the movement goal or motor plan into motor commands that go directly to the muscles, but also generate an efference copy that informs the forward model about the issued commands. The forward model then predicts the sensory consequences of the issued motor commands. Once sensory feedback from the body and the environment is provided, predicted and actual sensory consequences are compared. If they do not match, an error — the sensory prediction error — is detected. This provides state estimation to the controller that then updates the motor command (Izawa and Shadmehr, 2011). The sensory prediction error is regarded essential in distinguishing between body-generated from environmental changes.

Numerous studies on motor adaptation have presented evidence for this model framework. After having adapted to a visuomotor rotation or a force field and generated compensatory commands, the subject continues to produce the adapted commands, even when the perturbation is removed (after-effects, Fig. 19A). These self-generated errors give evidence that the internal model was altered. This error-based learning has been shown to scale with the size of the sensory prediction error.

Besides error-based learning, subjects can also learn purely based on reward or reinforcement, i.e., binary feedback about success or failure. The learning patterns based on reward versus error feedback have shown to be distinct. Reward was shown to affect exploratory features of motor variability. When reaching to an invisible target, variability increased as the probability of reward decreased (Pekny et al., 2015). Punishment and reward have distinct effects on the learning process during a visuomotor rotation task. While punishment (negative feedback) increases the speed of learning more than reward (positive feedback), reward leads to longer retention of the learnt movement (Galea et al., 2015). However, reward-based learning generalizes only locally, suggesting that error and reward-based learning are distinct processes (Izawa and Shadmehr, 2011). A third type of learning is use-dependent learning referring to the observation that a movement tends to be similar to the previous movement (Diedrichsen et al., 2010; Verstynen and Sabes, 2011). For example, when repeatedly reaching to a sequence of targets, the future movements will be biased towards the direction of the previous movement. Use-
dependent learning is considered as a model-free process, i.e., it does not depend on the state of the internal models.

9.4. Variability, noise and redundancy

The adage ‘repetition without repetition’, expresses that no movement is ever the same; even highly trained athletes never repeat the same movement twice. A pervasive assumption, reflected in the initial definition of learning, is that variability is undesired and needs to be reduced or eliminated (Guthrie, 1935). Overt variability arises from noise and fluctuations that are inherent to all sensory and motor processes, starting from the planning to the execution stage of movements, and is intrinsic to any interaction with objects in a variable environment (Faisal et al., 2008; Sternad, 2018). Hence, skill cannot, and probably should not, completely suppress noise. Rather, it should “make noise matter less”, i.e., have little or no effect on task success (Sternad et al., 2014). In fact, variability can also be beneficial for skill learning, as for example when exploring possible strategies to learn a novel task. This initial exploration followed by exploitation has been recognized as an essential transition in the acquisition of a skill (Dhawale et al., 2017).

This overt variability is also the direct result of the redundancy of the complex neuromotor system, i.e., the neuromotor system has many more degrees of freedom than necessary to achieve a given task goal. A movement as common as leading a cup of coffee to one’s mouth can be achieved with many different hand and arm configurations. Given this many-to-one mapping between execution and task outcome, our sensorimotor system seems to exploit this opportunity and varies performance with every new attempt (Bernstein, 1967). This variability opens an avenue for a more differentiated analysis of motor control and can sensitively characterize individual performance and its change in the process of learning. Fig. 22 illustrates the simplified example how two execution variables can be combined in different ways and still achieve the same result. The curved surface represents the solution manifold that contains all executions that lead to zero error (Cusumano and Cesari, 2006; Latash et al., 2002; Müller and Sternad, 2009). Importantly, amongst the infinitely many performance variations, the central nervous system does not use all possible variants that achieve the task goal. Analysis of the distribution of selected strategies can shed light on the question which subset of strategies the central nervous system selects and why.

9.5. Stages of learning

Practicing a new sufficiently complex skill typically begins by exploring the space of possible solutions until a ‘ballpark’ or subset of solutions is identified that promise task success. This exploration stage is followed by a fine-tuning of the skill that reduces overt variability and improves accuracy of the desired task result. Beyond documenting this transition via outcome measures, quantitative analysis of the distribution of variability over repeated executions presents a sensitive way to characterize these stages. Based on the redundancy in the high-dimensional space of executions, performance improvements are characterized as an overlapping sequence of three distinct processes: increasing tolerance to perturbations and intrinsic noise, channelling of variability into task-irrelevant directions by covarying execution variables, and by attenuating the level of noise (Cohen and Sternad, 2009). Fig. 22 illustrates this evolution with three exemplary data sets on successive practice days; each data point represents one execution. Starting with a large dispersion on day 1, the variability decreases with practice and approaches the solution manifold, i.e., approaches zero-error solutions. Due to the redundancy, the data can be aligned with the solution manifold by covarying the execution variables so that they do not affect the result. In a final stage, noise is reduced, i.e., the dispersion is reduced. The sequential evolution of skill has also been described as a transition from an initially cognitive to an autonomous stage: initial identification and selection of a suitable strategy involves cognitive decision processes, while the later stage of tuning leads to an increasing automatization that requires less attention to performance details, hence shifting the emphasis to lower-level neural resources (Schneider and Shiffrin, 1977).

Learning implies that performance changes persist beyond the practice sessions, although typically with some decay in the longer term. This long-term stability is brought about not only during practice but also in an off-line consolidation process (Krakauer and Shadmehr, 2006). This consolidation process involves neural processes that happen after the practice has concluded. This effect has been particularly pronounced after a night’s sleep (Robertson et al., 2004).

9.6. Neuroplasticity underlying motor learning

With these many facets of motor learning, underlying neural processes also differ across the different types of motor learning tasks and stages of learning (Dayan and Cohen, 2011; Floyer-Lea and Matthews, 2005; Krakauer et al., 2019). The main tool used to study the neurophysiology of motor learning in human beings has been neuroimaging, via functional magnetic resonance imaging (fMRI), along with EEG. TMS is used to artificially stimulate neural populations, with concomitant impacts on behavior and learning. Most neuroimaging studies have examined simple tasks with relatively fast performance improvements as this facilitates observations in the laboratory. Further, studies in humans with EEG, fMRI and TMS have only been performed on relatively small-scale movements as larger movements involving multi-
joint or whole-body behaviors can create recording artifacts. Hence, studies on adaptation to visuomotor rotation and learning of finger sequences are predominant in human neurophysiology research. Sequence learning, specifically the serial reaction time task (SRTT), involves finger movements in response to stimuli. As learning a sequence of cues involves memorizing strings of items, this task has a strong cognitive component and its neural processes differ from coordination tasks (Robertson, 2007). Hence, here we focus on adaptation and skill learning that are predominantly sensorimotor behaviors.

Regardless of the specific experimental paradigm, the main cerebral areas engaged in motor learning are the primary motor cortex (M1), premotor cortex, primary somatosensory cortex (S1), supplementary motor area (SMA), and subcortical structures, particularly the cerebellum, basal ganglia, and thalamus; complementary contributions come from parietal and prefrontal areas (dorsolateral prefrontal cortex, DLPFC). The final common pathway for motor learning is, of course, the spinal cord, which also exhibits learning and plasticity entirely on its own. The known primary mechanisms of neuromodulation, as identified in rodents, monkeys and humans, are dendritic arborization, synaptogenesis, remapping of cortical areas, structural changes in grey and white matter, and changes in neural firing and functional connectivity. Important to note is that learning is not associated with a uniform increase or decrease of these neuroplastic processes, but rather it is a subtle re-weighting of processes across the evolution of a motor skill. Given this breadth of phenomena, some basic results are grouped around two main types of motor learning, adaptation and de novo skill acquisition, with a focus on results from humans.

9.6.1. Motor adaptation

A wide range of studies on adaptation paradigms agree that the cerebellum plays a central role (Cullen and Brooks, 2015; Sokolov et al., 2017). Individuals with cerebellar lesions exhibit profound deficits in adapting to visuomotor rotation, external force fields, and modified locomotor demands, the latter introduced by split treadmills (Martin et al., 1996; Morton and Bastian, 2006; Rabe et al., 2009). Early recalibration to such altered environments correlates with the initially large errors that drive rapid plasticity in the cerebellum (Criscimagna-Hemminger et al., 2010). As improvements progress, more gradual plasticity is observed in the deep cerebellar nuclei. For example, activity in lobule VI was strongly correlated with the amount of savings measured in the retest session (Debas et al., 2010). Stimulation via direct transcranial stimulation of M1 that did not affect the rate of adaptation, but showed a marked increase on the retention of the adapted behavior (Galea et al., 2011). Similarly, disruption of the posterior parietal cortex using tTMS had no effect on the early phase of adaptation but decreased the amount of adaptation reached at steady state. Similarly, the basal ganglia seem to play a subordinate role in adaptation (Della-Maggiore et al., 2004). Patients with basal ganglia dysfunction, such as Parkinson and Huntington disease, retain their ability to adapt and only show some decrements in the after-effects (Marinelli et al., 2009).

Additional connectivity analyses of fMRI data have characterized large-scale functional reorganization, integrating nonadjacent brain regions into functional cortico–cortico and cortico-cerebellar networks. Two main networks have been identified: a M1–premotor–parietal–cerebellar circuit that reduces activity as learning progresses. The second is a posterior parietal–premotor circuit that shows increasing activity and correlates with behavioral gains (Hikosaka et al., 2002).

9.6.2. Skill acquisition

Considerably fewer studies have examined the acquisition of novel coordinative skills in humans, due to practical considerations mentioned above. A notable difference between adaptation and de novo learning is the role of the motor cortex that is central for skill acquisition. One explanation is that skill requires learning of new muscle activations by the motor cortex which directly sends motor commands to the muscles to generate forces. The output of the motor cortex to the spinal cord and its communication with an intact corticospinal tract determines the degree of achievable skill.

Most studies on skilled motor coordination used animal models where more invasive imaging and recording techniques can be leveraged. Numerous rodent studies have examined neuroplasticity in the motor cortex in the context of training a prehensile skill, typically reaching through an aperture to grab a pellet. These studies have yielded important information about structural and functional changes in the motor cortex showing that motor skill relies on the formation and selective maintenance of new synapses (Peters et al., 2017). In rats, it could be shown that dendritic branching in increased with training duration, although this arborization is a combination of forming new spines while eliminating old ones. Specifically, spines that present inputs from inhibitory neurons are eliminated, suggesting that changes in skill is associated with a balanced evolution of excitatory and inhibitory processes (Chen et al., 2015). However, any increase in connectivity of the neural network is temporary and returns back to normal without loss of skill. Hence, increases in strength of individual synapses are not the mechanism of storage, rather it is the pattern of connectivity within an ensemble of cells. It is indeed intuitive that new muscle combinations, or muscle synergies, for new motor skills emerge from the pattern of connectivity from the cortex to descending pathways (Overduin et al., 2012). Note though that corresponding assessment of the behavioral changes in rodents has largely been confined to coarse-grained outcome measures, such as successful completion of a grasp or locomotor velocity on a
treadmill. That is, the links between neural activity and the details of a learned behavior have yet to be established.

Invasive neurophysiology in non-human primates has focused on target-directed reaching movements. Traditional analysis has focused on control, not addressing learning, and used single neuron analyses (Cisek and Kalaska, 2005, Georgopoulos et al., 1982). This approach correlates intracortically measured spike trains of single neurons and correlates spike activity (analyzed in raster plots) with movement to identify what feature neurons may encode, e.g., velocity or direction (Fig. 23A). More recent work has turned to de novo skill learning, mostly in the context of brain-machine interfaces. In these studies, invasive recordings of populations of neurons have been collected and used to drive a cursor on a screen or an external robotic device (Ganguly and Carmena, 2009; Sadler et al., 2014; Serruya et al., 2002; Taylor et al., 2003). Thus, brain-computer interfaces causally relate neural firing to behavior. To understand the process of learning such new interfaces, a new analysis method has been developed that extended the single-neuron analyses to the analysis of neuronal populations (Churchland et al., 2012). Fig. 23B sketches population analysis, where the firing rates of recorded neurons are plotted against each other, creating a high-dimensional neural space. After dimensionality reduction analysis systematic patterns emerge in the form of neural trajectories. It was shown that when a new movement is learnt, this neural activity is confined to a smaller-dimensional subspace or a neural manifold (Gallego et al., 2020; Sadler et al., 2014). Recent studies addressed whether the learnt manifold presents limits for learning (Golub et al., 2018). Altering the mapping between neural activity and cursor output requires learning, i.e., changes in neural activity. Results showed that new task demands that required neural activity outside the manifold necessitated significantly longer practice than remapping of neural activity within the neural manifold (Oby et al., 2019).

Many neurophysiological studies on skill learning revealed significant remapping of the motor-cortical landscape with training. Such maps are obtained through systematic grid-like stimulation of the motor cortical area and measurement of the elicited muscle activity in the body. Training prehension tasks in rodents has been associated with an expansion of the territory for the trained effectors (Klein et al., 1998). One study on humans learning a motor sequence showed an increase in the size of motor maps and cortico-motor-neuronal excitability of the digits measured by TMS (Pascual-Leone et al., 1994). Critically, these maps are not permanent and longer-term retention is associated with a shrinking of the motor-cortical area and measurement of the elicited muscle activity. The newly learned information in the acquisition phase is thought to be processed and reactivated offline, such that memory traces become fixed through a cascade of events occurring at both the synaptic and the brain systems levels. This consolidation of motor skill representations is observed hours or days after the first training session, although the same offline phenomena may also occur rapidly between blocks of practice executed during the training session (Robertson, 2009). Consolidation is particularly effective after a night’s sleep (Doyon and Benali, 2005; Eichenlaub et al., 2020). EEG recordings (that have sufficient temporal resolution) have identified ‘sleep spindles’, i.e., bursts of neural firing at 11–16 Hz, that appear responsible for the integration of multiple learning-specific brain regions into synchronized oscillatory activity (Boutin and Doyon, 2020). This functional reorganization and consolidation of memory traces via synchronization is assumed to create cortical-subcortical networks involving the hippocampus, striatum, thalamus and motor-related cortical regions for motor memories.

9.6.3. Relation between learning and recovery

Given this rich literature on motor learning, a pressing question is to what degree lessons from healthy motor learning can inform rehabilitation to promote recovery of patients with neurological injury such as stroke (Krakauer, 2006). The fundamental mechanisms of neuroplastic processes in healthy motor learning—synaptogenesis, dendritic growth, structural and functional reorganization—are likely to be valid for individuals with neuro-

Fig. 23. Intracortical recording of neural activity and single neuron and population analysis. A: Neural recordings and raster plot. Using an electrode array individual neurons are recorded from the same movement and averaged over many repetitions. Spike density indicates activity that correlates with movement, shown as a kinematic trajectory. B: Trajectories in neural space. Recordings from dozens to hundreds of neurons are plotted against each other (n1, n2, n3, ... nm) to create a multidimensional neural space. Neural firing rates in a lower dimensional space manifest as trajectories. C: Neural manifold. The yellow surface illustrates the subspace of the neural trajectories. New mappings can be readily learnt if the new demands require neural activity within the manifold (grey dimension). Neural activity orthogonal to the manifold (blue arrow), requires significantly longer practice or cannot be learnt.
logical lesions. As reviewed above, after ischemic strokes rehabilitation of skilled hand function is accompanied by functional reorganization of the cortical map (Nudo et al., 1996). These results suggest that rehabilitative training can promote such recovery-related reorganization in the adjacent intact cortex. However, it remains unknown whether different locations or extents of lesions require specific types of motor training or training schedules. Nonetheless, in the recent decades knowledge of motor control has begun to be applied to characterize and treat motor deficits after hemiparesis. Rehabilitation practice is starting to embrace techniques that promote formation of appropriate representations beyond sheer repetition of movements (Hanlon, 1996). For example, variable and interleaved practice of different task variations is more likely to lead to persistent changes that also generalize beyond the practiced behavior, which is evidently the ultimate goal of rehabilitative practice.

But differences between learning and re-learning are also clear. Longitudinal studies revealed that recovery from stroke follows a series of fairly stereotypical stages over the first 6 months post-stroke, irrespective of the kind of therapeutic intervention (Kwakkel et al., 2006). An initial process of spontaneous recovery is expressed in the first four weeks post-stroke and then tapers off over subsequent months. Several likely mechanisms for this spontaneous recovery are restitution of the ischemic penumbra, resolution of diaschisis, and functional reorganization. Naturally, some aspects of brain reorganization are likely to be unique to the specific brain injury. Another important fact is that stroke typically involves damage to descending white matter pathways, such as the corticospinal tract, and motor recovery is related to the degree of integrity of these tracts (Lin et al., 2019). However, it is unknown how to best access these processes and harvest neuroscientific insights for therapeutic improvements.

A remaining puzzle is to dissociate between true recovery and compensation. True recovery means that undamaged brain regions are recruited to generate commands to the same muscles as were used before the injury. Compensation, in contrast, is the use of alternative muscles to accomplish the task goal. These modifications rely on the redundancy on the musculo-skeletal system. Further, a substantial contribution to post-stroke dysfunction, specifically in upper limb hemiparesis, is weakness, synergies, lack of dexterity, and spasticity. To date, there is no understanding how different learning processes relate to these behavioral changes and it remains often unclear whether learning per se is impaired or whether it is muscle weakness or spasticity that prevent the expression of voluntary actions. To resolve these open questions requires more quantitative movement analysis in addition to the outcome measures from the prevalent clinical scales such as the Fugl-Meyer scale or the Wolf-Motor Function Test. There is a lack of longitudinal studies that could address how their recovery relates to anatomy and physiology of the motor system. Given this sparsity of knowledge, the application of motor learning principles to therapy remains greatly in need for more scientific study.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: [U.Z. received grants from the German Ministry of Education and Research (BMBF), European Research Council (ERC), German Research Foundation (DFG), Janssen Pharmaceuticals NV and Takeda Pharmaceutical Company Ltd., and consulting fees from Bayer Vital GmbH, Pfizer GmbH and CorTec GmbH, all not related to this work. D.S received grants from NIH-R01-CRCNS-N5120579: Collaborative Research: Neural basis of motor expertise and NSF-MX-1825942: Collaborative Research: Learning to control dynamically complex objects. None of the other authors report any Conflict of Interest.]

Acknowledgements

Some sections of the article are similar to the corresponding sections in the Handbook of Clinical Neurophysiology, Volume 1 (Hallett, 2003), since this is intended to be an update of the Handbook.

References


