Estimating cognitive load during self-regulation of brain activity and neurofeedback with therapeutic brain-computer interfaces

Robert Bauer[1,2]* **and Alireza Gharabaghi**[1,2]*

[1] Division of Functional and Restorative Neurosurgery and Division of Translational Neurosurgery, Department of Neurosurgery, Eberhard Karls University Tuebingen, Tuebingen, Germany
[2] Neuroprosthetics Research Group, Werner Reichardt Centre for Integrative Neuroscience, Eberhard Karls University, Tuebingen, Germany

Neurofeedback (NFB) training with brain-computer interfaces (BCIs) is currently being studied in a variety of neurological and neuropsychiatric conditions in an aim to reduce disorder-specific symptoms. For this purpose, a range of classification algorithms has been explored to identify different brain states. These neural states, e.g., self-regulated brain activity vs. rest, are separated by setting a threshold parameter. Measures such as the maximum classification accuracy (CA) have been introduced to evaluate the performance of these algorithms. Interestingly enough, precisely these measures are often used to estimate the subject's ability to perform brain self-regulation. This is surprising, given that the goal of improving the tool that differentiates between brain states is different from the aim of optimizing NFB for the subject performing brain self-regulation. For the latter, knowledge about mental resources and work load is essential in order to adapt the difficulty of the intervention accordingly. In this context, we apply an analytical method and provide empirical data to determine the zone of proximal development (ZPD) as a measure of a subject's cognitive resources and the instructional efficacy of NFB. This approach is based on a reconsideration of item-response theory (IRT) and cognitive load theory for instructional design, and combines them with the CA curve to provide a measure of BCI performance.

Keywords: neurofeedback, cognitive load theory, zone of proximal development, workload, instructional design, brain-computer interface

INTRODUCTION

Brain-computer interfaces (BCIs) support reinforcement learning of brain self-regulation by feedback and reward. While assistive BCIs aim to replace lost functions by controlling external devices (Yanagisawa et al., 2011; Hochberg et al., 2012; Collinger et al., 2013; Wang et al., 2013), the ultimate goal of restorative or therapeutic approaches is to improve specific functions by neurofeedback (NFB) training, e.g., hand and arm control following a stroke (Ang et al., 2010; Shindo et al., 2011; Buch et al., 2012; Ramos-Murguialday et al., 2013; Gharabaghi et al., 2014a,b). The fundamental approach of NFB is based on the idea that physiological signals during restful waking (Mantini et al., 2007; Albert et al., 2009; De Vico Fallani et al., 2011) are contrasted to the signals during the task condition, using classification algorithms to weight the respective features (Theodoridis and Koutroumbas, 2009). In this regard, restorative BCI is similar to assistive BCI. However, unlike assistive BCI approaches, which select features on the basis of their ability to maximally contrast the two states, the feature space for restorative BCI and NFB training is constrained in accordance with the specific treatment rationale. In stroke rehabilitation, for example, the feature space might be restricted

to power in the β-range (15–30 Hz), since decreased movement-related desynchronization in this frequency range is related to the amount of motor impairment after the insult (Rossiter et al., 2014). During restorative BCI training, the power of the frequency band is therefore estimated, and the desired modulation of this feature space is reinforced using appropriate visual, auditory or haptic feedback (Gharabaghi et al., 2014a,b). The feature weights are deliberately constrained during these interventions and the modality of feedback is designed to maximize the reinforcing effect of NFB (Sherlin et al., 2011; Vukelić et al., 2014). By contrast, during assistive BCI, the feature weights are calculated so as to allow maximal separation (Blankertz et al., 2008; Theodoridis and Koutroumbas, 2009) and the classification output is used for communication or robotic control (Wolpaw et al., 2002). While the primary goal for assistive BCI is accuracy and speed, the main goal for restorative BCI is reinforcement and learning. A theoretical difference therefore exists between the self-regulation of brain activity and the classification algorithm (Wood et al., 2014). However, although there are several ways of measuring the performance of an assistive BCI (Thomas et al., 2013; Thompson et al., 2013), similar appropriate measures for restorative BCI

and NFB are currently not available. The most common measure for BCI is classification accuracy (CA). While the magnitude of CA has been used to estimate subject's ability to perform brain self-regulation, this interpretation currently lacks theoretical foundation (Blankertz et al., 2010; Buch et al., 2012; Hammer et al., 2012). What is more, there is no consensus regarding what approaches are appropriate for disentangling the performance of subject and classifier from each other, nor is there any theory as to how they are connected with each other.

By integrating classification theory (Theodoridis and Koutroumbas, 2009) with item response theory (Safrit et al., 1989), we describe how the relationship between the classification algorithm and the ability for self-regulation can be understood. In addition, on the basis of the theory of cognitive load for instructional design (Sweller, 1994; Schnotz and Kürschner, 2007), we will describe how the CA can be interpreted within the framework of NFB training. Our argument is based on the fact that it is possible to make an off-line calculation of the positive rates for different classifiers and thresholds. We will argue that the true positive and the false positive curve provide information about the subject's ability and his/her performance when support is provided. Moreover, since the shape of CA depends on the difference between true and false positive rate (FPR), we propose that it contains information about the subject's zone of proximal development (ZPD). Therefore, on the basis of the theory of cognitive learning, the ZPD may serve as an indirect measure of the subject's cognitive resources (Allal and Ducrey, 2000; Schnotz and Kürschner, 2007).

In this respect, the goal of this paper is to provide a measurement theory for subjects' abilities and ZPD during NFB training. We support this theory by mathematical models and by evidence from empirical data.

EMPIRICAL DATASET

Exemplary data is based on two right-handed, healthy subjects, one female (age 19) and one male (age 31), who presented different abilities for brain self-regulation. They each performed 75 trials of cued motor imagery. The trial structure consisted of consecutive preparatory (2 s), motor imagery (6 s) and rest (6 s) phases, each of which was initiated by a specific auditory cue. Electroencephalography (EEG) was recorded at 64 channels in accordance with the 10–20 system with Brain Products amplifiers and analyzed offline with custom-written scripts and Fieldtrip in Matlab (Oostenveld et al., 2011) according to the following steps. Data was down-sampled to 200 Hz and band-pass filtered between 14 and 26 Hz using a Butterworth filter. Wavelet transformation was used to perform a time-frequency analysis for time steps of 50 ms for the power in the β-range (15–25 Hz) over sensorimotor regions (FC3, C3 and CP3). For each trial, the power at each time point was normalized by z-scoring based on the mean and standard deviation of the power distribution in the rest and preparatory phase. Both subjects gave written, informed consent prior to participation. The study was approved by the local ethics committee.

LINKING SUBJECT'S ABILITY FOR BRAIN SELF-REGULATION WITH THE CLASSIFICATION PERFORMANCE

In the following section, we will propose a link between the ability for brain self-regulation, as estimated by the item function, with the classifier performance, as estimated by the rate function. This integration will enable us to apply off-line analysis of the positive rate across different thresholds to determine the subjects' ability for brain self-regulation.

RATE FUNCTIONS

When applying NFB for therapeutic purposes, a two-class separation of brain states is usually performed, i.e., rest vs. learned self-regulation of brain activity. Due to the fact that most of the classifiers used in NFB are based on supervised learning algorithms employing linear discriminant analysis, the sensitivity and specificity of the classifier can be calculated relatively easily. The sensitivity informs us how often the classifier detects sufficient self-regulation while the subject is performing brain self-regulation (true positive rate or TPR). The specificity informs us how often the classifier detects rest while the subject is performing insufficient brain self-regulation (true negative rate or TNR). Since the probabilities of each conditional classification must add up to 1 within each class, the false negative rate (FNR) is equal to 1-TPR, and the FPR is equal to 1-TNR. CA is based on the average of TPR and TNR.
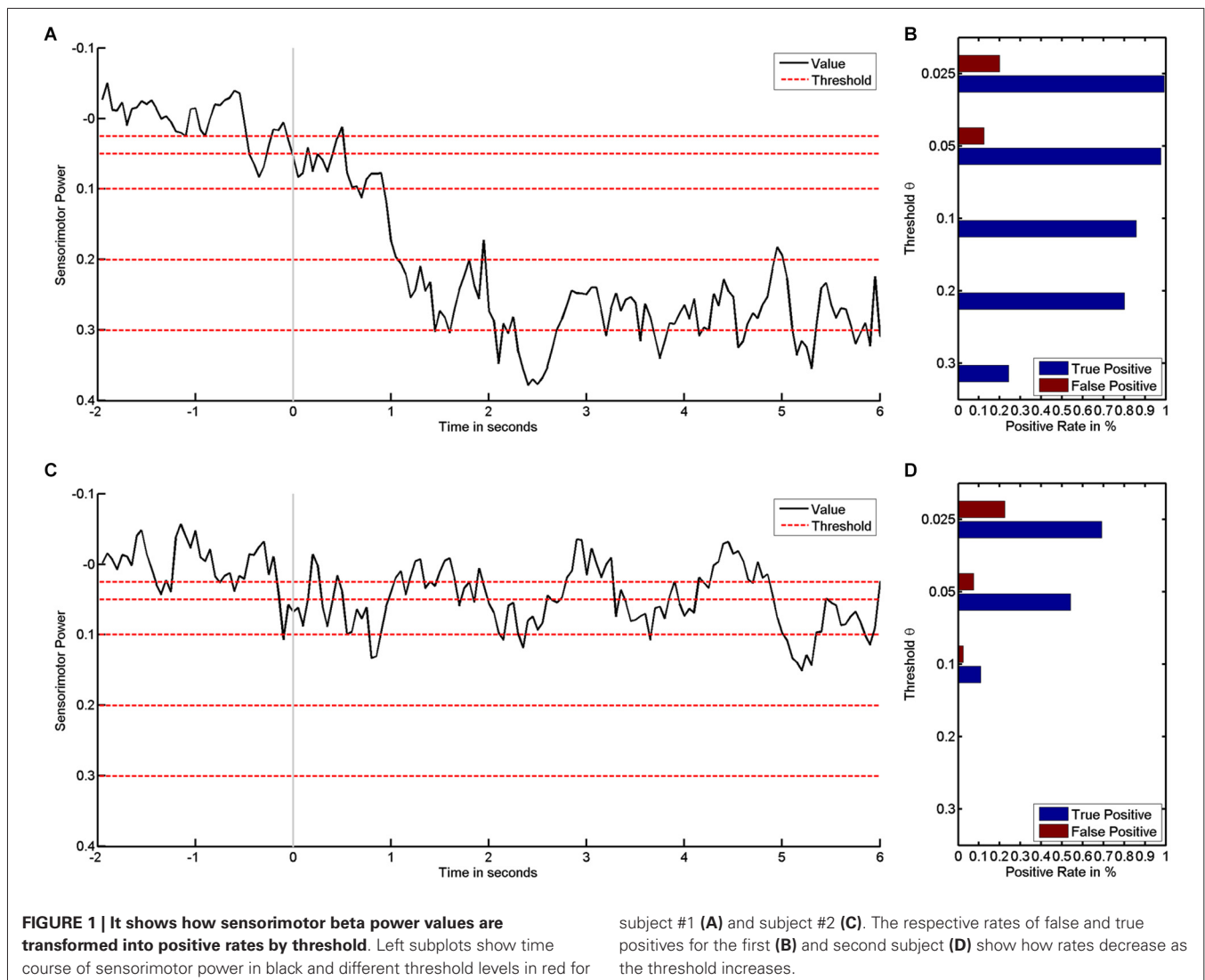
$$CA = \frac{(TPR + TNR)}{2}$$

THRESHOLD-BASED RATE FUNCTIONS

These rates are functions of the threshold θ (Theodoridis and Koutroumbas, 2009). During the training, the threshold θ usually remains fixed, and the rates therefore also remain fixed. However, provided that the electrophysiological signals have been recorded and stored, the positive rate can be calculated offline after the training for any threshold. We exemplify this by the empirical dataset (see **Figure 1**): the higher—i.e., the more challenging—the threshold, the stronger the desynchronization must be if it is to be classified as positive (see **Figure 1A**). The examples also reveal how subjects vary in their ability for brain self-regulation, e.g., sensorimotor beta-band desynchronization. The first subject shows stronger desynchronization and is thus able to reach more challenging thresholds (see **Figure 1A**) than the second subject, who has less pronounced brain self-regulation (see **Figure 1C**).

In parallel, this data enables us to characterize the classifier performance: detecting "positive" during the motor imagery phase (second 0 to 6) is a true positive, whereas "positive" during the preparatory phase (second −2 to 0) is a false positive. The average rate of positives for each phase is a suitable measure for characterizing a classifier's performance, i.e., true and FPRs are expressed as probabilities in the range between 0 and 1 (see **Figures 1B/D**). The first subject has higher TPRs for all thresholds (see **Figure 1B**) than the second subject (see **Figure 1D**).

For most classifiers, the rate functions result in sigmoidal rate curves. We show this sigmoidal shape for the TPR

FIGURE 1 | It shows how sensorimotor beta power values are transformed into positive rates by threshold. Left subplots show time course of sensorimotor power in black and different threshold levels in red for subject #1 **(A)** and subject #2 **(C)**. The respective rates of false and true positives for the first **(B)** and second subject **(D)** show how rates decrease as the threshold increases.

of the empirical dataset (see **Figure 2A**). The higher the thresholds, the more the probability of success decreases in a logistic fashion. Accordingly, the location of the first subject's true positive curve is further to the right, indicating a generally higher success rate. However, the shape of the respective curves for the first and the second subject are highly similar.
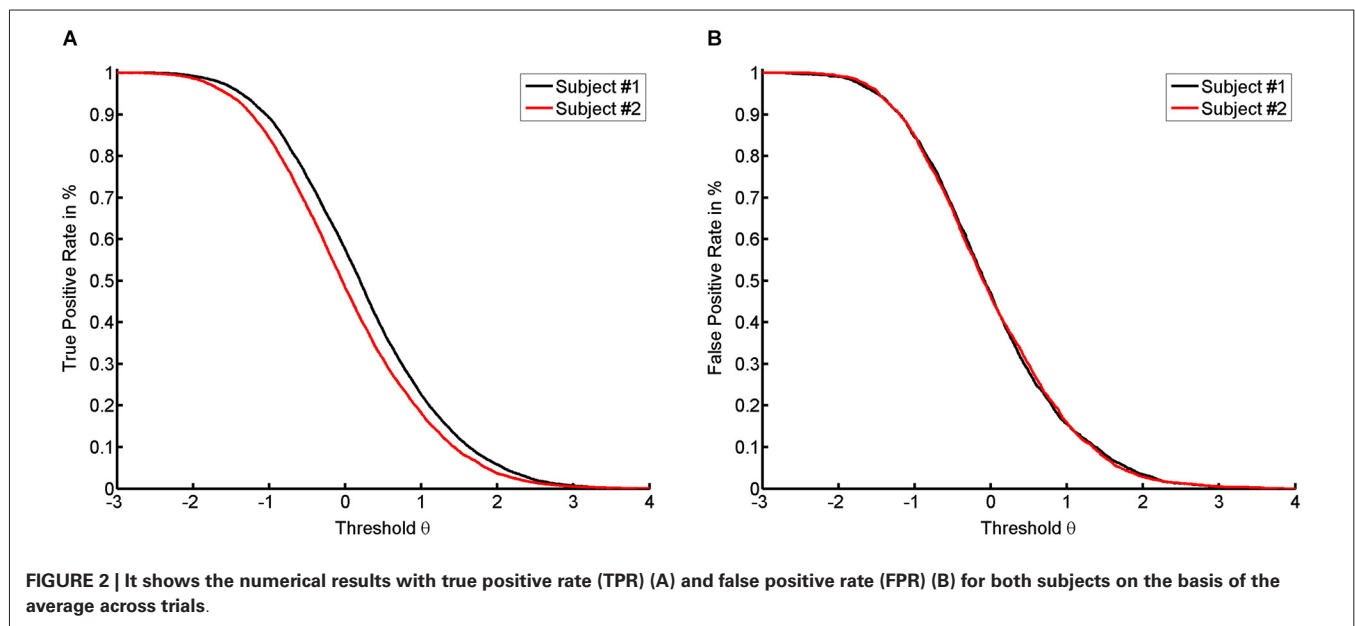
ITEM FUNCTIONS

The *rate functions* described resemble the *item functions* used in the psychometric item-response theory (IRT). In various fields of research, such as in assessment psychology (De Champlain, 2010) or motor behavior research (Safrit et al., 1989), the parameters of the item functions are usually estimated across datasets of several subjects and items. This enables us to quantify the respective variability of subject's ability and task difficulty necessary for fitting algorithms. If, for example a mathematical test battery is distributed to a school class, the marginal success rates enable us to estimate the difficulty of

a specific test and the ability of a single subject. Generally speaking, students with a higher mathematical ability achieve a higher success rate, and easier tests should result in higher average success rates. This information about the relative ranks based on success rates can be used for parameter estimation. More specifically, the shape of these success curves can be approximated by a two-parameter logistic model (2PLM) using the following function (Safrit et al., 1989; De Champlain, 2010):

$$P = \frac{1}{1 + e^{-D(\alpha - \theta)}}$$

In this function, *e* is Euler's number and D is the slope of the curve. The parameter θ, generally known as the threshold parameter in rate functions, now represents the difficulty of the task. Ability α and difficulty θ are located on the same axis. They can therefore be measured in one dimension. The location of the curve depends on the difference between the difficulty of a task and the subject's ability α. The shape of the curve depends on the slope parameter D, and a value of ~1.7 would result in

FIGURE 2 | It shows the numerical results with true positive rate (TPR) (A) and false positive rate (FPR) (B) for both subjects on the basis of the average across trials.

an approximate fit for a normal distribution (De Champlain, 2010). If the slope is identical for all curves, all item function are parallel and the difficulty level θ is the only changeable parameter.

SIMILARITY OF RATE AND ITEM FUNCTIONS

For dichotomous items, the probability of success is modeled as a function of the difficulty of the item and the ability of the subject. Assuming that the latter is stable, the difficulty of an item then depends on the design of the task. The combination of NFB task, i.e., the classification algorithm, the trial structure, the cues and any instructions or extraneous aspects constitute the phrasing of such an item. By way of example: in NFB training, our aim is to differentiate between sufficient and insufficient modulation of brain activity. If we were to use a questionnaire instead of EEG and BCI, we might phrase an item: "Are you currently performing sufficient brain self-regulation?" to which the possible answers would be "yes" and "no". However, since the decision about "yes" or "no" is based on physiological recordings during the task, a *post hoc* reassessment for different thresholds is possible. This recalculation enables us to apply virtually the same items over a range of difficulties. In addition, threshold-based recalculation modifies only one aspect of the "item", namely the difficulty parameter; an aspect that lies in one clearly defined dimension. These properties (uni-dimensionality, off-line analysis) allow for an interpretation of the parameters that describe the positive rate curves on the threshold dimension within the framework of the item response theory.

INTERPRETATION OF CURVE PARAMETERS
TRUE POSITIVE CURVE REPORTS ON ABILITY

Due to the fact that the measurement of the subject's ability for brain self-regulation can be performed *post hoc*, NFB training is comparable to an action like videotaping a sports exercise such

as long jump. The data acquired during the task can then be used later to estimate several aspects of the performance. As in sports training, the compound ability can be divided into sub-sets (e.g., sprinting, take-off, and landing in long jump). In this example, the coach would be ill-advised to reward any jump independent of the actual performance. Therefore, specificity matters.

NFB training has the ability to provide this specificity by selecting the appropriate features and classification algorithms. In addition to determining the threshold to be passed by the event-related desynchronisation (ERD), additional features might include the speed of the power dip in the first two seconds or continuity of desynchronization. This highlights the fact that the reinforced features must be carefully selected for their respective clinical or rehabilitative purpose.

In this respect, it is also important to note that functional improvement is a combination of several abilities and preconditions, of which for instance, brain self-regulation of sensorimotor beta-rhythms is only one example. Others, such as reaching out and holding a certain position with the upper limb, as assessed in the Fugl-Meyer assessment (Deakin et al., 2003), or interacting with an object, as assessed in the Broetz hand assessment (Brötz et al., 2014), necessitate the involvement of parieto-frontal circuits for motor planning (Andersen and Cui, 2009) and sensorimotor circuits for execution (Chouinard and Paus, 2006). Along these lines, stroke survivors who train to modulate the activity of the primary motor cortex (Kaiser et al., 2011; Kilavik et al., 2012) show improvements in this ability only if the fronto-parietal integrity is preserved (Buch et al., 2012). Training such ability of brain-self-regulation may therefore be related both directly and indirectly to the respective function, e.g., moving the upper extremity. However, improving brain self-regulation does not necessarily lead to functional improvements, since these may also depend on abilities and preconditions that are not influenced by the NFB

training. Improvements are therefore required with regard to the clinical efficacy of such a training (Ramos-Murguialday et al., 2013; Ang et al., 2014) by researching the feedback modality (Gomez-Rodriguez et al., 2011) or using it in combination with simultaneous cortical stimulation (Gharabaghi et al., 2014a). Screening examinations might also be necessary to determine the eligibility of subjects for a specific intervention (Stinear et al., 2012; Burke Quinlan et al., 2014; Vukelić et al., 2014). In addition, the validity of functional assessment scores requires re-evaluation in the light of biomarkers of sub-clinical improvement.

We therefore conclude that the location of the TPR can be interpreted as a subject's ability for brain self-regulation, regardless of the potential influence of this ability on a specific function. In this respect, the location of the true positive curve—mathematically speaking, the point of maximal slope and halfway between success and failure—provides information about the subject's ability to perform the task which is defined by the features and the classifier.

FALSE POSITIVE CURVE REPORTS ON ATTEMPT
According to our previous example, a long jump coach would be ill-advised not to reward any jump. To be more precise, for reasons of motivation, even attempts should sometimes be rewarded, or support is required to transform an attempt into a success. In the case of NFB, specificity and sensitivity also have to be balanced according to their importance for learning. If the task remains identical, such a balance can only be achieved by changing the threshold. Decreasing the threshold increases the number of false positives (see **Figure 2B**). Since the classifier normalizes to rest, there is no apparent difference in the location of the FPR between the two subjects (see **Figure 2B**). This indicates that the subjects have the same opportunity to try to perform the task. This theory is supported by the following line of argument. In this context, "support" or "help" can be formalized by assuming that the subject with the current level of ability α is unable to perform the task at the given level of difficulty θ, whereas providing help will lead to success. If we then detect a success, this will be a "false positive" result, since the subject's current ability is too low for him/her to actually succeed. If no help is provided, the success achieved will be a "true positive" result. This approach will lead to a range of thresholds which are defined by two limits. The lower limit will be marked by the most difficult task that the subject can perform when help is provided. The upper limit is defined by the most difficult task that can be performed by the subject without help (see **Figure 3A**). Once the subject no longer benefits from help, e.g., due to overly high intrinsic or extrinsic load, he can no longer benefit from the training.

SHAPE OF CLASSIFICATION ACCURACY SHEDS LIGHT ON THE ZONE OF PROXIMAL DEVELOPMENT
The range between the most difficult tasks that can be achieved by the subject with and without help, respectively, can be defined as the ZPD. Cognitive load theory argues that mental load can be divided into three categories: intrinsic load, extrinsic load and germane load (Jong, 2009). Intrinsic load resembles the difficulty of the task and is mainly caused by the element interactivity of the

task. Extraneous load is mainly caused by irrelevant information. Germane load is caused by the construction and subsequent automation of schemas, i.e., learning. Lower difficulty results in a reduction of intrinsic load while extrinsic load will increase, since the instructional material now contains irrelevant information. If a task is too easy (i.e., $\theta \ll \alpha$) or too difficult (i.e., $\theta \gg \alpha$) for a given ability, the extrinsic or the intrinsic load of the task would be too high.

For every given level of difficulty and ability for a task, there is therefore a ZPD, where learning is possible (Schnotz and Kürschner, 2007). The cognitive load theory thus provides a feasible explanation as to why the boundaries of ZPD are characterized by TPR and FPR (Allal and Ducrey, 2000; see **Figure 3B**). A second line of argumentation considers the likelihood of reward. Since a subject cannot discern a reward with identical qualities, the only way of differentiating between a true and a false reward is to determine the relative probability of their occurrence. The difference between the true and FPR might therefore be a good approximation of the difference with respect to the informational content of the two reward rates. Although more elaborate measures might be better suited to this divergence (MacKay, 2003), the most straightforward approach would consider the magnitude and the shape of ZPD as estimated by a linear transformation of CA in accordance with the following equations:
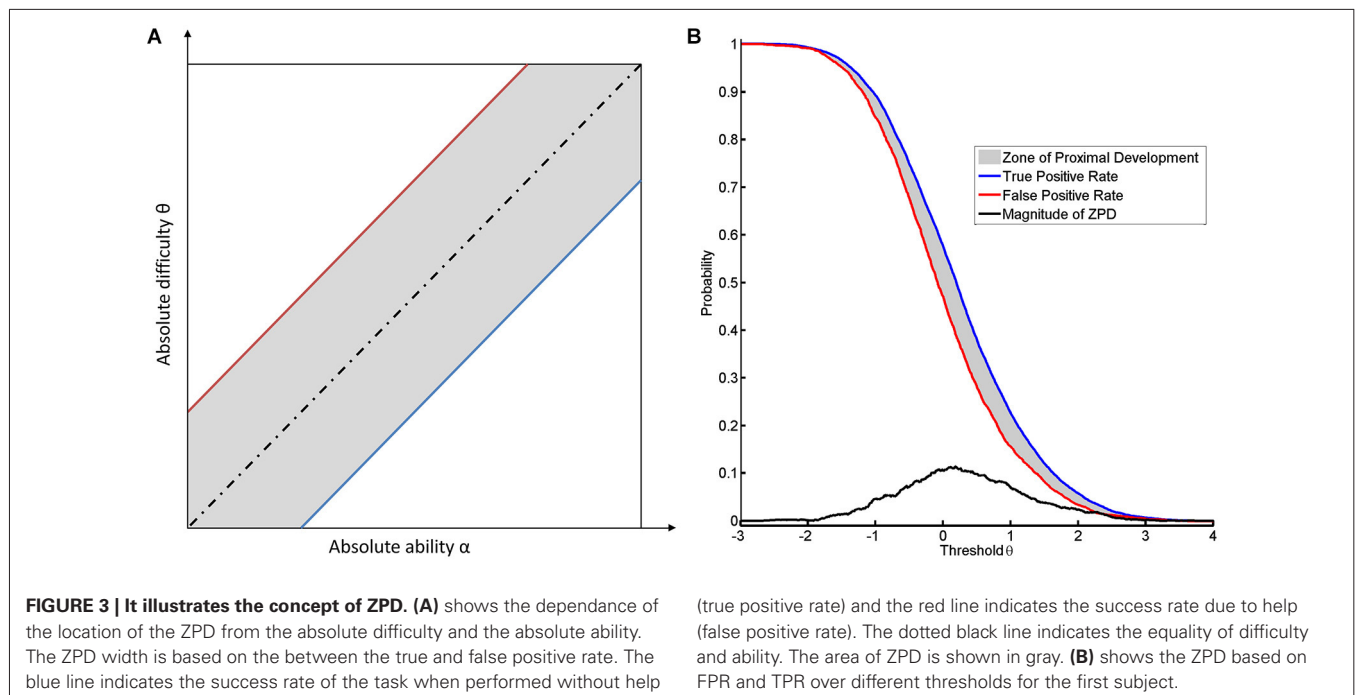
$$CA = \frac{TPR + TNR}{2}$$
$$ZPD = TPR - FPR$$
$$TNR = 1 - FPR$$
$$CA = \frac{TPR + (1 - FPR)}{2} = \frac{TPR - FPR + 1}{2}$$
$$(2 * CA) - 1 = TPR - FPR$$
$$ZPD = (2 * CA) - 1$$

CONCLUSION AND OUTLOOK
In the sections above, we have shown how the true and false positives rates of brain self-regulation can be interpreted within the framework of NFB. We have demonstrated that there is a natural relationship between classification of rate functions and item response functions. We have revealed the connection between applying a threshold to ERD signals and estimating the ability to perform ERD. In this respect, the true positive curve provides information as to the brain's ability to perform brain self-regulation in a NFB task. In addition, we showed that not only can the false positive curve provide information about attempts to perform the task but it can also set the lower limit of the ZPD. Below, we will illustrate how the ZPD, in its capacity as a transformation of CA, can support the instructional design of NFB interventions.

CONCLUSION REGARDING CLASSIFICATION ACCURACY
The ZPD can be used to compare different classification algorithms. In BCI approaches, classification algorithms are often trained to maximize CA. This can result in a peaky but narrow ZPD (see **Figure 4A**) instead of a flat but broad ZPD (see **Figure 4B**), although the area of ZPD is equal in both cases. A

FIGURE 3 | It illustrates the concept of ZPD. (A) shows the dependance of the location of the ZPD from the absolute difficulty and the absolute ability. The ZPD width is based on the between the true and false positive rate. The blue line indicates the success rate of the task when performed without help (true positive rate) and the red line indicates the success rate due to help (false positive rate). The dotted black line indicates the equality of difficulty and ability. The area of ZPD is shown in gray. **(B)** shows the ZPD based on FPR and TPR over different thresholds for the first subject.

more broadly shaped ZPD indicates that learning can occur over a larger range of thresholds, whereas a peaky shape means that maximal help is available only for a narrow range of thresholds. This being the case, slight misalignments might have significant adverse effects. The shape of CA may therefore serve as a measure to evaluate whether or not a NFB task is instructionally effective. While the best general instructional efficacy is obviously achieved by NFB training with a high *and* broad ZPD, interpreting the shape enables us to apply tailored approaches. These might be more effective with regard to instructional needs for specific subjects and environments. A broad ZPD might be more robust for home-based training with low availability of supervision and the possibility of noisy measurements. A peaky ZPD might be more suitable for environments where professionals can perform alignments, i.e., adapt the classification algorithm or correct noisy measurements. Shaping the ZPD can thus support instructional design of NFB interventions. The approach presented here will also be applicable to classification algorithms resulting in non-normal distributions where TPR and TNR are calculated numerically (see **Figure 3B**), since the interpretation is also supported by non-parametric IRT-models provided that TPR and FPR are monotonic functions (Mokken and Lewis, 1982; Rost, 2004).

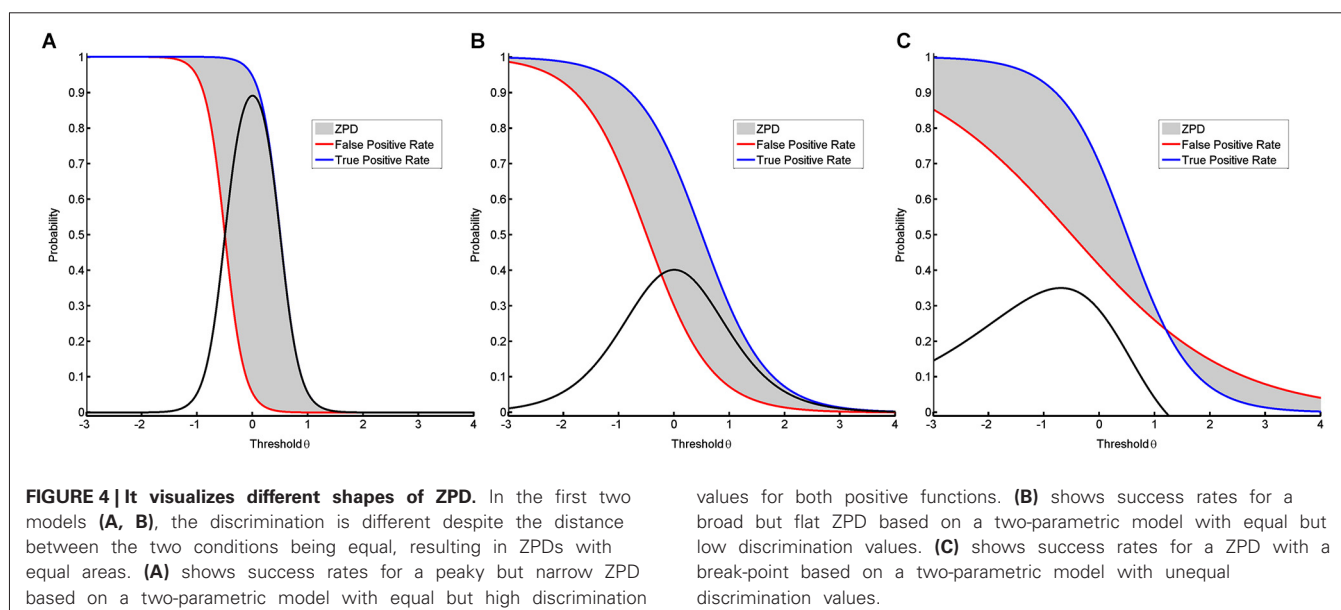### RELATIONSHIP TO ALTERNATIVE MEASURES OF PERFORMANCE AND FEEDBACK

CA is by far the most widely reported measure of performance for BCIs and can be used for both synchronized (e.g., cued) and self-paced interventions (Thomas et al., 2013). However, for some clinical applications, additional measures of performance that are regarded as relevant for the treatment goal have been developed. These include the latency to

movement onset or the maximum consecutive movement time in stroke rehabilitation (Ramos-Murguialday et al., 2013) and the path efficiency for a high degree of freedom prosthetic control in tetraplegia (Collinger et al., 2013). How does the ZPD now relate to these alternative performance measures?

These alternative measures may sometimes be translated into one of the basic measures used for calculating the ZPD, e.g., the average movement rate could also be understood as a TPR. However, such measures are much more liable to contain unique information about additional abilities that are required for the given task to be performed, such as already mentioned in the paragraph on the interpretation of the true positive curve.

Since learning is conceptually linked to the accuracy of feedback, we propose that a NFB task should be characterized by its instructional efficacy with regard to the action to be trained. This instructional efficacy is characterized by the feedback curves. In this context, the CA changes as soon as the coupling of the feedback to the action changes. The shape of the ZPD will therefore be useful for the instructional design of the intervention and tends to be independent of other task-specific measures. If, for example, the subject receives feedback to alternative actions, any improvement in these actions will be caused by the task's instructional design. In this respect, a ZPD, e.g., for latency of movement onset or path efficiency, may also exist. Estimating the ZPD for these measures would be similar to the approach illustrated above.

It should be borne in mind that the theory presented here is based on the classical binary feedback, the distance between feedback and no feedback being one bit of information, i.e., reward (Ortega and Braun, 2010). Alternative approaches such as

FIGURE 4 | It visualizes different shapes of ZPD. In the first two models **(A, B)**, the discrimination is different despite the distance between the two conditions being equal, resulting in ZPDs with equal areas. **(A)** shows success rates for a peaky but narrow ZPD based on a two-parametric model with equal but high discrimination values for both positive functions. **(B)** shows success rates for a broad but flat ZPD based on a two-parametric model with equal but low discrimination values. **(C)** shows success rates for a ZPD with a break-point based on a two-parametric model with unequal discrimination values.

continuous feedback (e.g., the frequency of an auditory signal) or psychophysical perception rules (e.g., the perception of the duration of binary feedback in a log-linear fashion) do, of course, affect the bit-rate and may thus increase the achievable speed of learning. However, since the ZPD is based on a single bit as a distance metric, adequate mapping of the ZPD for alternative feedback approaches will probably be mathematically demanding. In order to interpret the curves under such conditions, further research and specific transformations might be necessary. A system analytical perspective, where continuous feedback can be understood as a pattern of step functions, and a complex-valued ZPD might help to solve such aspects. Nonetheless, the real-valued ZPD based on the single bit of feedback will also remain a fundamental building block of such advanced approaches.

CONCLUSION WITH REGARD TO COGNITIVE RESOURCES

The ZPD may also act as a measure to compare different subjects with regard to their cognitive resources for a NFB task. If two subjects perform the very same NFB training, one might show a peaky and narrow ZPD (see **Figure 4A**) while the other has a broad and flat ZPD (see **Figure 4B**). Since in this case both curves indicate equal abilities, this difference in the respective shapes requires an alternative explanation. On the basis of the relevance of cognitive resources for the ZPD (Schnotz and Kürschner, 2007), we postulate that the shape of ZPD can also be applied to measure a subject's cognitive resources for coping with the mental load that occurs during a misalignment between ability and difficulty. Such an interpretation would, furthermore, permit a different view on the discussion about BCI illiteracy (Vidaurre and Blankertz, 2010). In particular, when the curves of TPR and FPR cross, they provide information about the specific break-points of that task. At this point, any support provided by the instructional design of the training will cease to be beneficial and will begin to be detrimental for the performance (see

Figure 4C). This would be indicated by a negative value for the ZPD.

However, these concepts require validation by future research. Measurements of cognitive resources are currently based on psychophysiological recordings (e.g., heart rate variability, blink rate, electrodermal response), which are highly variable and very difficult to generalize across task conditions (Cegarra and Chevalier, 2008; Novak et al., 2010). Motor imagery itself can also cause vegetative effects related to the imagined movement, e.g., subjects imagining running at 12 km/h had an increased heart rate and pulmonary ventilation as compared to walking at 5 km/h (Decety et al., 1991). Mental imagery might therefore affect psychophysiological biomarkers, masking the measurement of the mental effort unrelated to the imagery content. One alternative to psychophysiological measures is the application of self-rating questionnaires. However, from the subject's point of view, it is often not possible to distinguish between the intrinsic, extrinsic and germane load (Cegarra and Chevalier, 2008). What is more, many psychophysiological measures and questionnaires can be sampled only at a very low rate. For example, the low frequency part of the heart rate commences at 0.04 Hz (Malik et al., 1996), meaning that at least 25 s of clean data have to be recorded for adequate frequency resolution of the Fourier transformation. Furthermore, slow frequency fluctuations in the EEG (<0.1 Hz) can correlate with psychophysiological performance, but they require similarly large time windows. In addition, slow fluctuations in the EEG measurements appear to be highly masked by imagery-related fluctuations, e.g., movement-related cortical potentials (Shibasaki and Hallett, 2006). This is also an issue if higher frequency components of the EEG are used to estimate cognitive resources, e.g., in the gamma range (Grosse-Wentrup et al., 2011), since they need to be disentangled from pure motor-related fluctuations in the same frequency band (de Lange et al., 2008; Miller et al., 2012).

In this context, the shape of ZPD might prove useful for disentangling the multitude of interacting and complex psychophysiological measurements in challenging tasks. This perspective is in agreement with the understanding that a proper alignment of ability and difficulty will reduce mental effort (Schnotz and Kürschner, 2007). Future studies might focus on psychophysiological correlates of the shape of the ZPD. Furthermore, improving the instructional material should help to reduce extrinsic load. NFB training could similarly be supported by "instructions", e.g., by providing haptic feedback (Gomez-Rodriguez et al., 2011) or visual and auditory cueing (Heremans et al., 2009). Systematic research on the impact of these feedback modalities on the ZPD might provide insight on their utility in guiding instructional design.

ACKNOWLEDGMENTS

REFERENCES

Albert, N. B., Robertson, E. M., Mehta, P., and Miall, R. C. (2009). Resting state networks and memory consolidation. *Commun. Integr. Biol.* 2, 530–532. doi: 10.4161/cib.2.6.9612

Allal, L., and Ducrey, G. P. (2000). Assessment of—or in—the zone of proximal development. *Learn. Instr.* 10, 137–152. doi: 10.1016/s0959-4752(99)00025-0

Andersen, R. A., and Cui, H. (2009). Intention, action planning and decision making in parietal-frontal circuits. *Neuron* 63, 568–583. doi: 10.1016/j.neuron.2009.08.028

Ang, K. K., Guan, C., Chua, K. S. G., Ang, B. T., Kuah, C., Wang, C., et al. (2010). Clinical study of neurorehabilitation in stroke using EEG-based motor imagery brain-computer interface with robotic feedback. *Conf. Proc. IEEE Eng. Med. Biol. Soc.* 2010, 5549–5552. doi: 10.1109/iembs.2010.5626782

Ang, K. K., Guan, C., Phua, K. S., Wang, C., Zhou, L., Tang, K. Y., et al. (2014). Brain-computer interface-based robotic end effector system for wrist and hand rehabilitation: results of a three-armed randomized controlled trial for chronic stroke. *Front. Neuroeng.* 7:30. doi: 10.3389/fneng.2014.00030

Blankertz, B., Sannelli, C., Halder, S., Hammer, E. M., Kübler, A., Müller, K.-R., et al. (2010). Neurophysiological predictor of SMR-based BCI performance. *Neuroimage* 51, 1303–1309. doi: 10.1016/j.neuroimage.2010.03.022

Blankertz, B., Tomioka, R., Lemm, S., Kawanabe, M., and Muller, K. (2008). Optimizing spatial filters for robust EEG single-trial analysis. *IEEE Signal Process. Mag.* 25, 41–56. doi: 10.1109/msp.2008.4408441

Brötz, D., Del Grosso, N. A., Rea, M., Ramos-Murguialday, A., Soekadar, S. R., and Birbaumer, N. (2014). A new hand assessment instrument for severely affected stroke patients. *NeuroRehabilitation* 34, 409–427. doi: 10.3233/NRE-141063

Buch, E. R., Modir Shanechi, A., Fourkas, A. D., Weber, C., Birbaumer, N., and Cohen, L. G. (2012). Parietofrontal integrity determines neural modulation associated with grasping imagery after stroke. *Brain* 135, 596–614. doi: 10.1093/brain/awr331

Burke Quinlan, E., Dodakian, L., See, J., McKenzie, A., Le, V., Wojnowicz, M., et al. (2014). Neural function, injury and stroke subtype predict treatment gains after stroke. *Ann. Neurol.* 77, 132–145. doi: 10.1002/ana.24309

Cegarra, J., and Chevalier, A. (2008). The use of Tholos software for combining measures of mental workload: toward theoretical and methodological improvements. *Behav. Res. Methods* 40, 988–1000. doi: 10.3758/brm.40.4.988

Chouinard, P. A., and Paus, T. (2006). The primary motor and premotor areas of the human cerebral cortex. *Neurosci. Rev. J. Bringing Neurobiol. Neurol. Psychiatry* 12, 143–152. doi: 10.1177/1073858405284255

Collinger, J. L., Wodlinger, B., Downey, J. E., Wang, W., Tyler-Kabara, E. C., Weber, D. J., et al. (2013). High-performance neuroprosthetic control by an individual with tetraplegia. *Lancet* 381, 557–564. doi: 10.1016/S0140-6736(12)61816-9

Deakin, A., Hill, H., and Pomeroy, V. M. (2003). Rough guide to the fugl-meyer assessment. *Physiotherapy* 89, 751–763. doi: 10.1016/s0031-9406(05)60502-0

Decety, J., Jeannerod, M., Germain, M., and Pastene, J. (1991). Vegetative response during imagined movement is proportional to mental effort. *Behav. Brain Res.* 42, 1–5. doi: 10.1016/s0166-4328(05)80033-6

De Champlain, A. F. (2010). A primer on classical test theory and item response theory for assessments in medical education. *Med. Educ.* 44, 109–117. doi: 10.1111/j.1365-2923.2009.03425.x

de Lange, F. P., Jensen, O., Bauer, M., and Toni, I. (2008). Interactions between posterior gamma and frontal alpha/beta oscillations during imagined actions. *Front. Hum. Neurosci.* 2:7. doi: 10.3389/neuro.09.007.2008

De Vico Fallani, F., Vecchiato, G., Toppi, J., Astolfi, L., and Babiloni, F. (2011). Subject identification through standard EEG signals during resting states. *Conf. Proc. IEEE Eng. Med. Biol. Soc.* 2011, 2331–2333. doi: 10.1109/iembs.2011.6090652

Gharabaghi, A., Kraus, D., Leão, M. T., Spüler, M., Walter, A., Bogdan, M., et al. (2014a). Coupling brain-machine interfaces with cortical stimulation for brain-state dependent stimulation: enhancing motor cortex excitability for neurorehabilitation. *Front. Hum. Neurosci.* 8:122. doi: 10.3389/fnhum.2014.00122

Gharabaghi, A., Naros, G., Walter, A., Grimm, F., Schuermeyer, M., Roth, A., et al. (2014b). From assistance towards restoration with an implanted brain-computer interface based on epidural electrocorticography: a single case study. *Restor. Neurol. Neurosci.* 32, 517–525. doi: 10.3233/RNN-140387

Gomez-Rodriguez, M., Peters, J., Hill, J., Schölkopf, B., Gharabaghi, A., and Grosse-Wentrup, M. (2011). Closing the sensorimotor loop: haptic feedback facilitates decoding of motor imagery. *J. Neural Eng.* 8:036005. doi: 10.1088/1741-2560/8/3/036005

Grosse-Wentrup, M., Schölkopf, B., and Hill, J. (2011). Causal influence of gamma oscillations on the sensorimotor rhythm. *Neuroimage* 56, 837–842. doi: 10.1016/j.neuroimage.2010.04.265

Hammer, E. M., Halder, S., Blankertz, B., Sannelli, C., Dickhaus, T., Kleih, S., et al. (2012). Psychological predictors of SMR-BCI performance. *Biol. Psychol.* 89, 80–86. doi: 10.1016/j.biopsycho.2011.09.006

Heremans, E., Helsen, W. F., De Poel, H. J., Alaerts, K., Meyns, P., and Feys, P. (2009). Facilitation of motor imagery through movement-related cueing. *Brain Res.* 1278, 50–58. doi: 10.1016/j.brainres.2009.04.041

Hochberg, L. R., Bacher, D., Jarosiewicz, B., Masse, N. Y., Simeral, J. D., Vogel, J., et al. (2012). Reach and grasp by people with tetraplegia using a neurally controlled robotic arm. *Nature* 485, 372–375. doi: 10.1038/nature11076

Jong, T. (2009). Cognitive load theory, educational research and instructional design: some food for thought. *Instr. Sci.* 38, 105–134. doi: 10.1007/s11251-009-9110-0

Kaiser, V., Kreilinger, A., Müller-Putz, G. R., and Neuper, C. (2011). First steps toward a motor imagery based stroke BCI: new strategy to set up a classifier. *Front. Neurosci.* 5:86. doi: 10.3389/fnins.2011.00086

Kilavik, B. E., Zaepffel, M., Brovelli, A., MacKay, W. A., and Riehle, A. (2012). The ups and downs of beta oscillations in sensorimotor cortex. *Exp. Neurol.* 245, 15–26. doi: 10.1016/j.expneurol.2012.09.014

MacKay, D. J. C. (2003). *Information Theory, Inference and Learning Algorithms.* Cambridge, UK: Cambridge University Press.

Malik, M., Bigger, J. T., Camm, A. J., Kleiger, R. E., Malliani, A., Moss, A. J., et al. (1996). Heart rate variability. *Eur. Heart J.* 17, 354–381. doi: 10.1093/eurheartj/17.suppl_2.28

Mantini, D., Perrucci, M. G., Del Gratta, C., Romani, G. L., and Corbetta, M. (2007). Electrophysiological signatures of resting state networks in the human brain. *Proc. Natl. Acad. Sci. U S A* 104, 13170–13175. doi: 10.1073/pnas.0700668104

Miller, K. J., Hermes, D., Honey, C. J., Hebb, A. O., Ramsey, N. F., Knight, R. T., et al. (2012). Human motor cortical activity is selectively phase-entrained on underlying rhythms. *PLoS Comput. Biol.* 8:e1002655. doi: 10.1371/journal.pcbi.1002655

Mokken, R. J., and Lewis, C. (1982). A nonparametric approach to the analysis of dichotomous item responses. *Appl. Psychol. Meas.* 6, 417–430. doi: 10.1177/014662168200600404

Novak, D., Mihelj, M., and Munih, M. (2010). Psychophysiological responses to different levels of cognitive and physical workload in haptic interaction. *Robotica* 29, 367–374. doi: 10.1017/s0263574710000184

Oostenveld, R., Fries, P., Maris, E., and Schoffelen, J.-M. (2011). FieldTrip: open source software for advanced analysis of MEG, EEG and invasive electrophysiological data. *Comput. Intell. Neurosci.* 2011, 1–9. doi: 10.1155/2011/156869

Ortega, P. A., and Braun, D. A. (2010). A conversion between utility and information. *Proc. Third Conf. Artif. Gen. Intell.* 115–120. doi: 10.2991/agi.2010.10

Ramos-Murguialday, A., Broetz, D., Rea, M., Läer, L., Yilmaz, O., Brasil, F. L., et al. (2013). Brain-machine-interface in chronic stroke rehabilitation: a controlled study. *Ann. Neurol.* 74, 100–108. doi: 10.1002/ana.23879

Rossiter, H. E., Boudrias, M.-H., and Ward, N. S. (2014). Do movement-related beta oscillations change following stroke? *J. Neurophysiol.* 112, 2053–2058. doi: 10.1152/jn.00345.2014

Rost, J. (2004). *Lehrbuch Testtheorie - Testkonstruktion.* Bern: Huber.

Safrit, M. J., Cohen, A. S., and Costa, M. G. (1989). Item response theory and the measurement of motor behavior. *Res. Q. Exerc. Sport* 60, 325–335. doi: 10.1080/02701367.1989.10607459

Schnotz, W., and Kürschner, C. (2007). A reconsideration of cognitive load theory. *Educ. Psychol. Rev.* 19, 469–508. doi: 10.1007/s10648-007-9053-4

Sherlin, L. H., Arns, M., Lubar, J., Heinrich, H., Kerson, C., Strehl, U., et al. (2011). Neurofeedback and basic learning theory: implications for research and practice. *J. Neurother.* 15, 292–304. doi: 10.1080/10874208.2011.623089

Shibasaki, H., and Hallett, M. (2006). What is the bereitschaftspotential? *Clin. Neurophysiol. Off. J. Int. Fed. Clin. Neurophysiol.* 117, 2341–2356. doi: 10.1016/j.clinph.2006.04.025

Shindo, K., Kawashima, K., Ushiba, J., Ota, N., Ito, M., Ota, T., et al. (2011). Effects of neurofeedback training with an electroencephalogram-based brain-computer interface for hand paralysis in patients with chronic stroke: a preliminary case series study. *J. Rehabil. Med.* 43, 951–957. doi: 10.2340/16501977-0859

Stinear, C. M., Barber, P. A., Petoe, M., Anwar, S., and Byblow, W. D. (2012). The PREP algorithm predicts potential for upper limb recovery after stroke. *Brain* 135, 2527–2535. doi: 10.1093/brain/aws146

Sweller, J. (1994). Cognitive load theory, learning difficulty and instructional design. *Learn. Instr.* 4, 295–312. doi: 10.1016/0959-4752(94)90003-5

Theodoridis, S., and Koutroumbas, K. (2009). *Pattern Recognition.* Burlington, MA: Academic Press.

Thomas, E., Dyson, M., and Clerc, M. (2013). An analysis of performance evaluation for motor-imagery based BCI. *J. Neural Eng.* 10:031001. doi: 10.1088/1741-2560/10/3/031001

Thompson, D. E., Blain-Moraes, S., and Huggins, J. E. (2013). Performance assessment in brain-computer interface-based augmentative and alternative communication. *Biomed. Eng. Online* 12:43. doi: 10.1186/1475-925x-12-43

Vidaurre, C., and Blankertz, B. (2010). Towards a cure for BCI illiteracy. *Brain Topogr.* 23, 194–198. doi: 10.1007/s10548-009-0121-6

Vukelić, M., Bauer, R., Naros, G., Naros, I., Braun, C., and Gharabaghi, A. (2014). Lateralized alpha-band cortical networks regulate volitional modulation of beta-band sensorimotor oscillations. *Neuroimage* 87, 147–153. doi: 10.1016/j.neuroimage.2013.10.003

Wang, W., Collinger, J. L., Degenhart, A. D., Tyler-Kabara, E. C., Schwartz, A. B., Moran, D. W., et al. (2013). An electrocorticographic brain interface in an individual with tetraplegia. *PloS One* 8:e55344. doi: 10.1371/journal.pone.0055344

Wolpaw, J. R., Birbaumer, N., McFarland, D. J., Pfurtscheller, G., and Vaughan, T. M. (2002). Brain-computer interfaces for communication and control. *Clin. Neurophysiol. Off. J. Int. Fed. Clin. Neurophysiol.* 113, 767–791. doi: 10.1016/S1388-2457(02)00057-3

Wood, G., Kober, S. E., Witte, M., and Neuper, C. (2014). On the need to better specify the concept of "control" in brain-computer-interfaces/neurofeedback research. *Front. Syst. Neurosci.* 8:171. doi: 10.3389/fnsys.2014.00171

Yanagisawa, T., Hirata, M., Saitoh, Y., Goto, T., Kishima, H., Fukuma, R., et al. (2011). Real-time control of a prosthetic hand using human electrocorticography signals. *J. Neurosurg.* 114, 1715–1722. doi: 10.3171/2011.1.jns101421