Clinical Neurophysiology 127 (2016) 3033–3041

Contents lists available at ScienceDirect

Clinical Neurophysiology

journal homepage: www.elsevier.com/locate/clinph

What is the optimal task difficulty for reinforcement learning of brain self-regulation?

Robert Bauer *, Mathias Vukelić, Alireza Gharabaghi *

Division of Functional and Restorative Neurosurgery, and Centre for Integrative Neuroscience, Eberhard Karls University Tuebingen, Germany

ARTICLE INFO

Article history:
Accepted 19 June 2016
Available online 25 June 2016

Keywords:
Beta modulation
Volitional control
Brain–computer interface
Brain–machine interface
Brain–robot interface

HIGHLIGHTS

- Empirical assessment of the incentive to learn brain self-regulation during operant conditioning.
- Learning incentive and classification accuracy peak at different task difficulties.
- Specificity is more important for learned brain self-regulation than sensitivity of neurofeedback.

ABSTRACT

Objective: The balance between action and reward during neurofeedback may influence reinforcement learning of brain self-regulation.
Methods: Eleven healthy volunteers participated in three runs of motor imagery-based brain–machine interface feedback where a robot passively opened the hand contingent to β-band modulation. For each run, the β-desynchronization threshold to initiate the hand robot movement increased in difficulty (low, moderate, and demanding). In this context, the incentive to learn was estimated by the change of reward per action, operationalized as the change in reward duration per movement onset.
Results: Variance analysis revealed a significant interaction between threshold difficulty and the relationship between reward duration and number of movement onsets ($p < 0.001$), indicating a negative learning incentive for low difficulty, but a positive learning incentive for moderate and demanding runs. Exploration of different thresholds in the same data set indicated that the learning incentive peaked at higher thresholds than the threshold which resulted in maximum classification accuracy.
Conclusion: Specificity is more important than sensitivity of neurofeedback for reinforcement learning of brain self-regulation.
Significance: Learning efficiency requires adequate challenge by neurofeedback interventions.

1. Introduction

Neurofeedback and brain-interface technology are being increasingly applied in fields of research aiming to restore upper-limb functionality in stroke survivors. Greater gains are currently being achieved by subacute (Pichiorri et al., 2015) than by chronic patients (Ang et al., 2014). On the basis of the neurophysiological correlates of motor imagery (Kaiser et al., 2011) and motor cortex excitability (Takemi et al., 2013; Kraus et al., 2016a), such as modulation of β-power (15–30 Hz), these devices may provide an effective backdoor to the motor system (Sharma 2006; Bauer et al., 2015), particularly when the subject receives contingent proprioceptive feedback with robotic rehabilitation technology (Gomez-Rodriguez et al., 2011; Vukelić et al., 2014; Vukelić and Gharabaghi, 2015).

When it comes to these restorative brain–robot interfaces (BRI), the classification of brain states is often *linear* (Theodoridis and Koutroumbas, 2009), i.e. based on thresholding, and *synchronous* (Thomas et al., 2013), i.e. based on instructive cues to initiate and stop modulation of brain activity. The feedback is usually *contingent*, i.e. linked to the maintenance of the desired neurophysiological state, and based on two classes such as rest and motor imagery. Within such a framework, a BRI for motor rehabilitation would, for example, rely on starting and stopping finger or arm extension by a robotic orthosis on the basis of the classification

* Corresponding authors at: Division of Functional and Restorative Neurosurgery, Eberhard Karls University, Otfried-Mueller-Str. 45, 72076 Tuebingen, Germany. Fax: +49 7071 29 25104.
E-mail addresses: robert.bauer@cin.uni-tuebingen.de (R. Bauer), alireza.gharabaghi@uni-tuebingen.de (A. Gharabaghi).

http://dx.doi.org/10.1016/j.clinph.2016.06.016

output (Brauchle et al., 2015; Naros and Gharabaghi, 2015). Linear and synchronous classifiers allow the straightforward calculation of performance metrics such as the true positive rate (TPR) or true negative rate (TNR), which equal the sensitivity and specificity of the classifier, respectively. Common performance metrics (Sokolova and Lapalme, 2009; Thomas et al., 2013; Thompson et al., 2013) suggest that accuracy and speed of the classifier are important goals. However, restorative BRIs can also be understood as a form of neurofeedback training, i.e. they should induce the operant conditioning of a specific brain state or modulation range. In this respect, motivation and reinforcement learning are further and perhaps even more important goals (Nijboer et al., 2010; Hammer et al., 2012; Lorenz et al., 2014).

1.1. Classification accuracy

Balanced classification accuracy (CA), i.e. the average of TPR and TNR, is the most common measure for assessing the performance of classification algorithms (Thompson et al., 2013). For classical *assistive* BRIs, which follow the goal to replace lost functions and control external devices, maximizing classification accuracy is considered to be optimal, and the threshold (θ) is selected accordingly (Theodoridis and Koutroumbas, 2009). This rationale is also implicitly assumed to be valid for *restorative* BRIs, which aim at neurofeedback training and use-dependent plasticity. Within this framework, a low classification accuracy is believed to result in decreased signal-to-noise ratio of the feedback which may, in turn, increase cognitive effort and reduce learning efficiency (Clark, 2006). From a learning perspective, however, the classification accuracy indicates whether the reward is specific and sensitive. We explored, therefore, a mathematical simulation based on a Bayesian reinforcement learning model to find out which thresholds are optimal for learning (Bauer and Gharabaghi, 2015b). This model revealed that learning occurred earliest at the threshold of maximum classification accuracy (θ_{CA}). The mathematical modeling indicated, moreover, that operant conditioning can be optimized when an adaptation strategy for threshold selection is applied in the course of the training (Bauer and Gharabaghi, 2015b). Such an adaptation strategy would need to change the classifier threshold, i.e., difficulty level, of the feedback device to challenge the participant in the course of the training. Moreover, the provided feedback should retain its specificity and reward trained actions rather than punish false ones (Bauer and Gharabaghi, 2015b; Naros and Gharabaghi, 2015).

1.2. Zone of proximal development (ZPD)

This information-theoretical understanding, i.e. when learning is achieved by processing the information content of the reward, is supported and extended from the perspective of cognitive load theory, where learning can occur at several difficulty levels as well. The range of difficulty levels, where the challenge and ability of a subject can be brought into alignment and where the subjects still have sufficient cognitive resources for the learning effort, is referred to as the ZPD (Schnotz and Kürschner, 2007). In that regard, the range of difficulties where a subject may learn from a neurofeedback task is not limited to the point of $\theta_{CA.}$ Instead, learning can occur over a range of thresholds (θ), i.e. difficulty levels. We have, therefore, suggested to interpret the classification accuracy curve over thresholds (see Fig. 1), where a classification accuracy above 50% is achieved, as indicative of the ZPD (Bauer and Gharabaghi, 2015a), a perspective which is in line with the Bayesian reinforcement learning model (Bauer and Gharabaghi, 2015b).

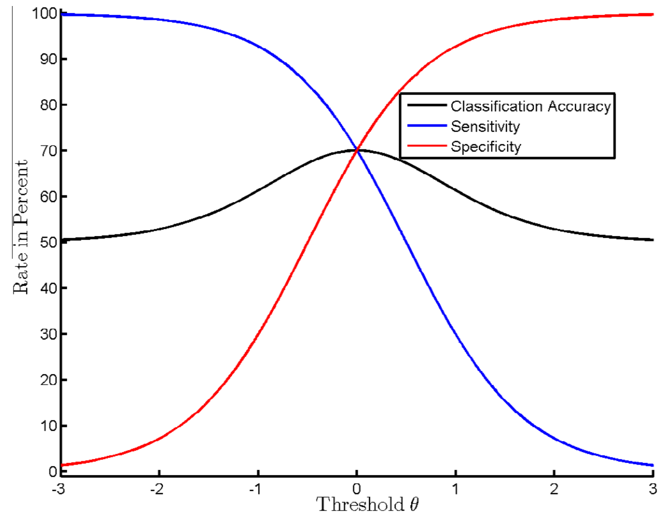

Fig. 1. Performance measures of a neurofeedback task. This figure exemplifies the evolution of the true positive rate or sensitivity (blue trace), the true negative rate or specificity (red trace) and the balanced classification accuracy (black trace) for different thresholds. The classification accuracy peaks at a distinct threshold (θ_{CA}).

1.3. Learning incentive

The information content of the feedback, i.e. whether the reward is specific and sensitive, indicates the potential but not the motivation for learning. In order to assess the instructional efficiency of different thresholds within the ZPD, i.e. their influence on the motivation to learn, a combination of cognitive load and expectancy-value theory would be necessary (Wigfield and Eccles, 2000; Schnotz and Kürschner, 2007; Sherlin et al., 2011; Lotte et al., 2013). More specifically: A low difficulty level corresponds to a high true positive rate of the classifier and thus to a high reward rate (see Fig. 1). Furthermore, the reward rate depends not only on the difficulty θ, but also on the subject's ability. The examiner can alter the reward rate only by changing the threshold. The subject, however, can alter the reward rate by increasing his/her ability, i.e. by learning, which in turn takes effort. Applying the expectancy-value theory of motivation (Wigfield and Eccles, 2000), the probability of actually investing this effort for learning would be proportional to the expected reward. Furthermore, learning requires an opportunity (O) to learn. In this context, we suggest that, within the framework of synchronous BRIs, every occasion on which the trained action is initiated can be interpreted as a learning opportunity. In this way, the learning incentive (I) can be expressed as proportional to the change in a subject's reward rate (R) per discrete repetition of the action (O):

$$I \sim \frac{\Delta R}{\Delta O}$$

To increase the incentive for learning in BRI tasks, the threshold θ_I, which maximizes $\Delta R/\Delta O$, should therefore be determined. Please note that this threshold (θ_I) is not necessarily identical to the threshold (θ_{CA}) which maximizes classification accuracy.

1.4. Operationalization

In the present study, the development of a criterion for detecting and optimizing θ_I was based on the following rationale. Due to the contingent feedback and the cued trial structure of a synchronous restorative BRI, the feedback returning to the subject in the course of a trial exhibits distinct patterns (see