Brain-computer interface performance development over 30 hours of patient training with an auditory event-related potential protocol

Michael Tangermann[1], Alexander de Ranitz[1], Soz Raheem[1] & Mariacristina Musso[2]

[1] Radboud University, Donders Institute for Brain, Cognition and Behaviour, Machine Learning and Neural Computing, Data-Driven Neurotechnology Lab, Nijmegen, the Netherlands

[2] German Center for Neurodegenerative Diseases (DZNE), Bonn, Germany

DONDERS INSTITUTE

INTRODUCTION

Background

BCI-supported rehabilitation after stroke is conducted over **multiple training sessions**. As shown for motor imagery protocols, patients can learn to modulate their brain activity over the course of the training by activating relevant brain networks differently. For protocols using event-related potentials (ERP), such longitudinal user learning scenarios are sparse.

Challenges

For **ERP protocols**, and specifically for stroke patients, it is unclear, **if patients can learn to improve the decoding performance**, and how it possibly develops over multiple sessions. Analyzing the performance development is challenging, as (i) the task conditions and difficulty may not be constant over time (to keep the patient challenged) and act as confounders, (ii) ceiling effects may prevent from observing an improvement over time, and (iii) patient training protocols require flexibility w.r.t. session durations. Overall, we would like to **understand better the side conditions under which a performance improvement can be observed** longitudinally.

Approach

We analyzed the epoch-based binary classification accuracy of **approx. 30 hours of EEG data each for 10 stroke patients with aphasia** who underwent an auditory, word-based ERP-BCI language rehabilitation protocol[1]. In a posthoc analysis, we evaluated the **learning curves** for three different classification strategies over time.

METHODS

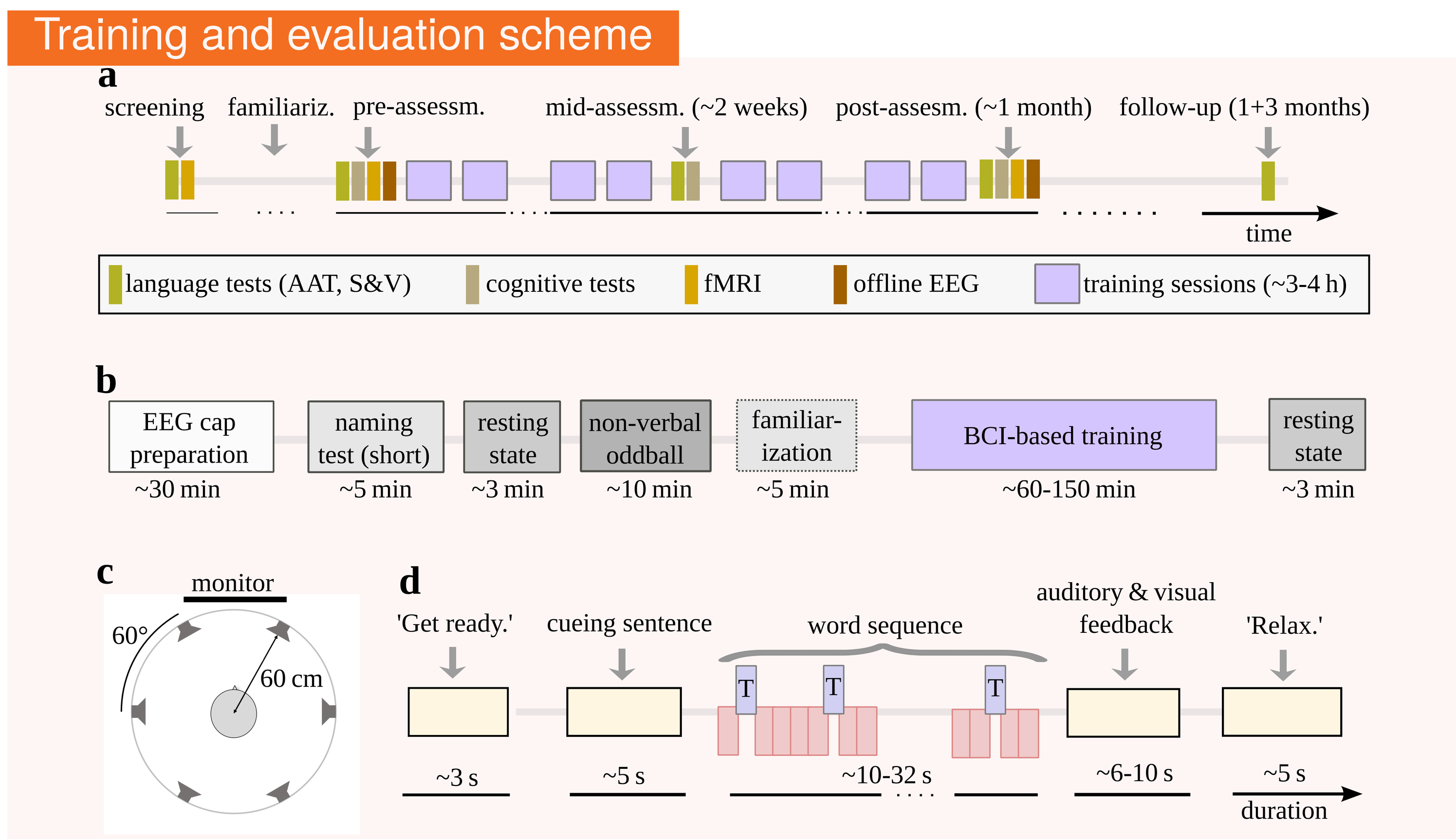
LONGITUDINAL PATIENT DATA

Dataset

Our study used an existing dataset: Ten patients, aged 38-76 with with different levels of stroke-induced aphasia had participated in a supervised, auditory, ERP-BCI feedback trainining for language rehabilitation[1].

Per patient, 31-channel EEG recordings of approx. 30 hours of closed-loop online feedback training (blue boxes) were collected during **14 to 28 sessions, each containing between 2940 and 15144 sequential word presentations**, i.e. epochs:

Training and evaluation scheme



Task, conditions and feedback

Following the AMUSE paradigm[5], patients were asked to focus attention on one target word and ignore five non-target words, which were played repeatedly in a pseudo-random order. One trial contained up to 90 word presentations (events/epochs).

Word stimuli were played either via 6 loudspeakers (**condition 6D**) or mono via headphones (**condition HP**) at an **stimulus onset asynchrony (SOA) ranging between 250 ms and 1000 ms**. Mono presentation and shorter SOA were used to control the task difficulty individually.

After each trial, a **patient received feedback** about how well the (correct) target word could be decoded based on an ERP analysis using an LDA classifier with shrinkage regularization. Producing words was not required.

POSTHOC DATA ANALYSIS

Preprocessing and classification

The raw 31-channel EEG data was band-pass filtered between 0.5 and 16 Hz. Per word stimulus, one epoch was extracted. For 18 short intervals and all channels, the average amplitude was extracted, leading to a feature vector of shape 1x558. 738 outlier epochs were removed due to excessive amplitudes.

Epoch-wise binary classification performance was obtained based on 5-fold cross-validation using a **block-Toeplitz linear discriminant analysis (btLDA)**[6].

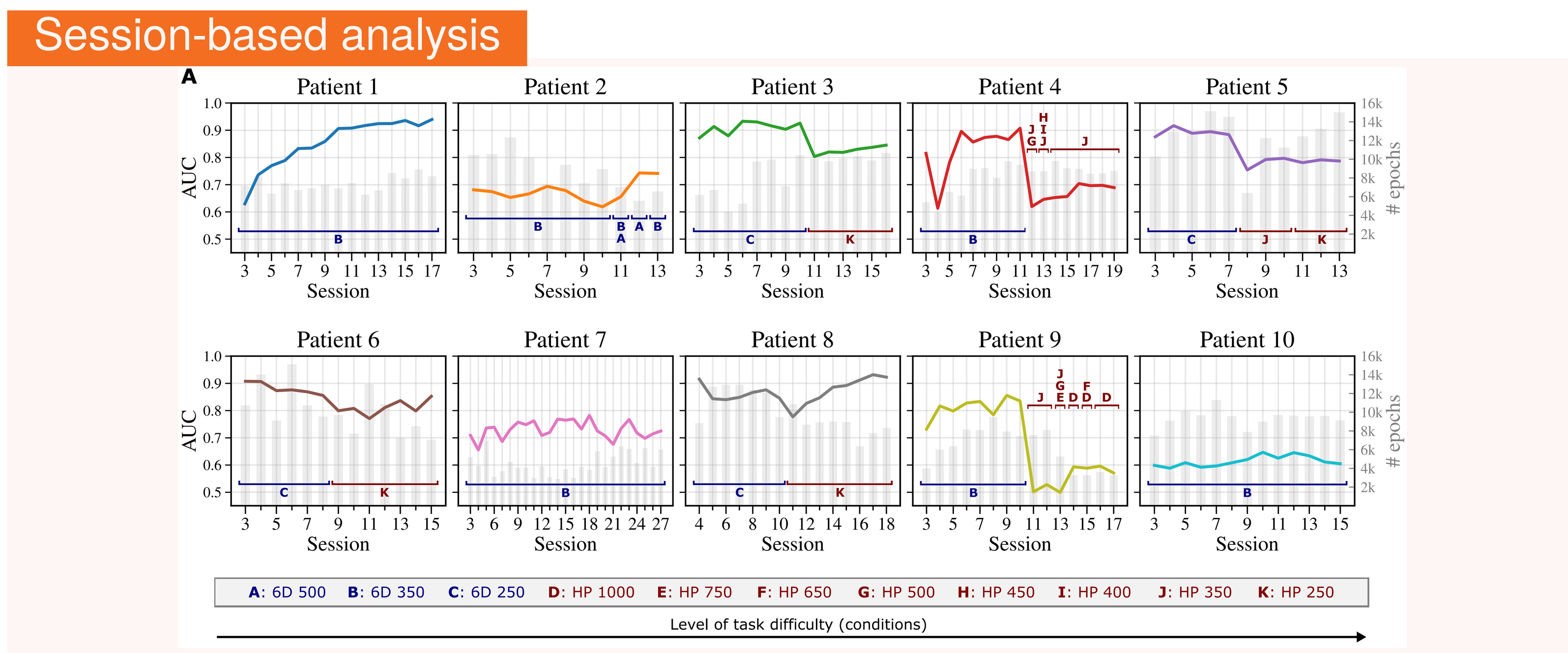
To be robust regarding class imbalances and biases, the performance values are reported using the area under the receiver-operating characteristic curve (ROC-AUC).

Adaptive classification over time

1. **Session-based** analysis: respects the original session structure, but the varying number of epochs per session biases the performance level. This analysis contains a mix of task conditions only for the rare cases, when conditions were changed within a session.
2. **Moving window-based** analysis: uses a fixed number of 5000 epochs and a stride of 1000 epochs for unbiased performance levels. A window may span across sessions and task conditions — in these cases, the reported performance may be lower than in windows spanning only a single task or session.
3. **Stable conditions** analysis: windows (5000 epochs, stride of 1000 epochs) span chronologically across a continuous period of a constant condition, which can contain multiple sessions. Avoids systematic performance breakdowns due to changed conditions.

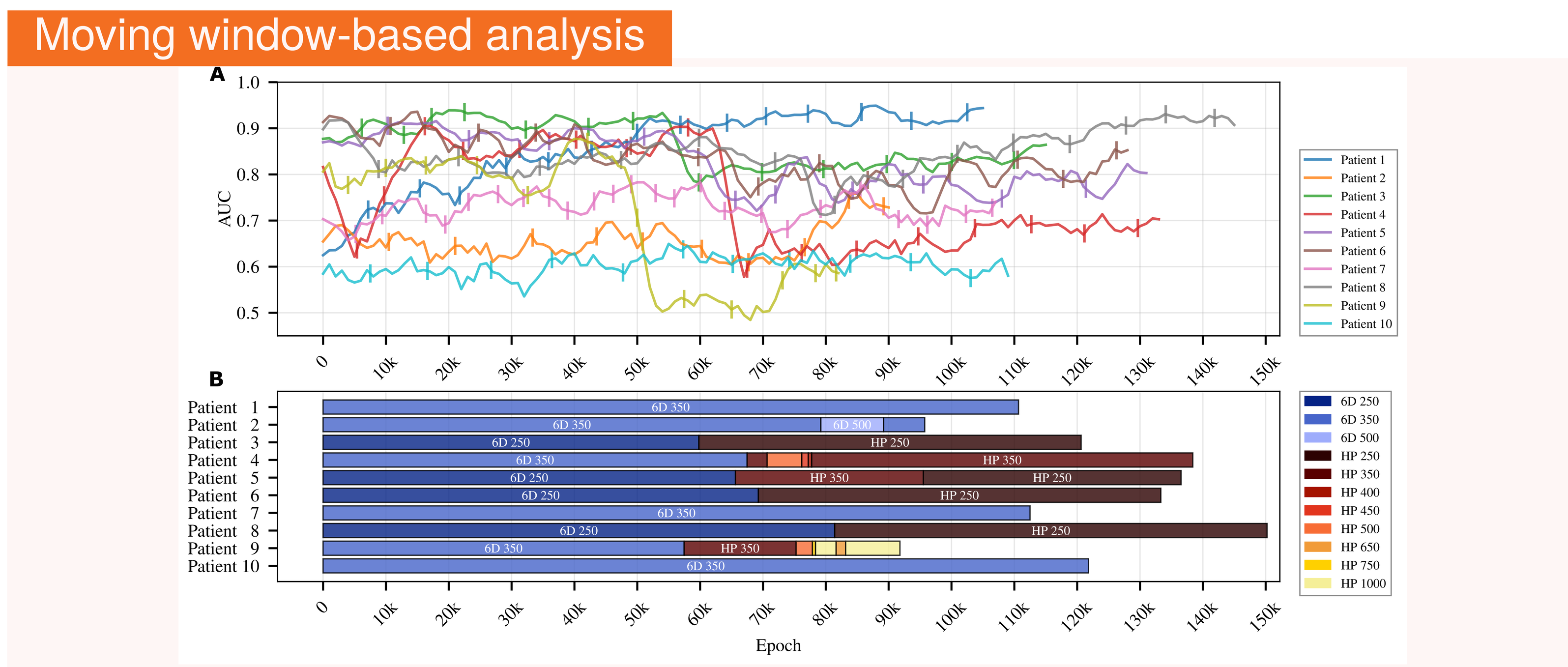
RESULTS

Session-based analysis



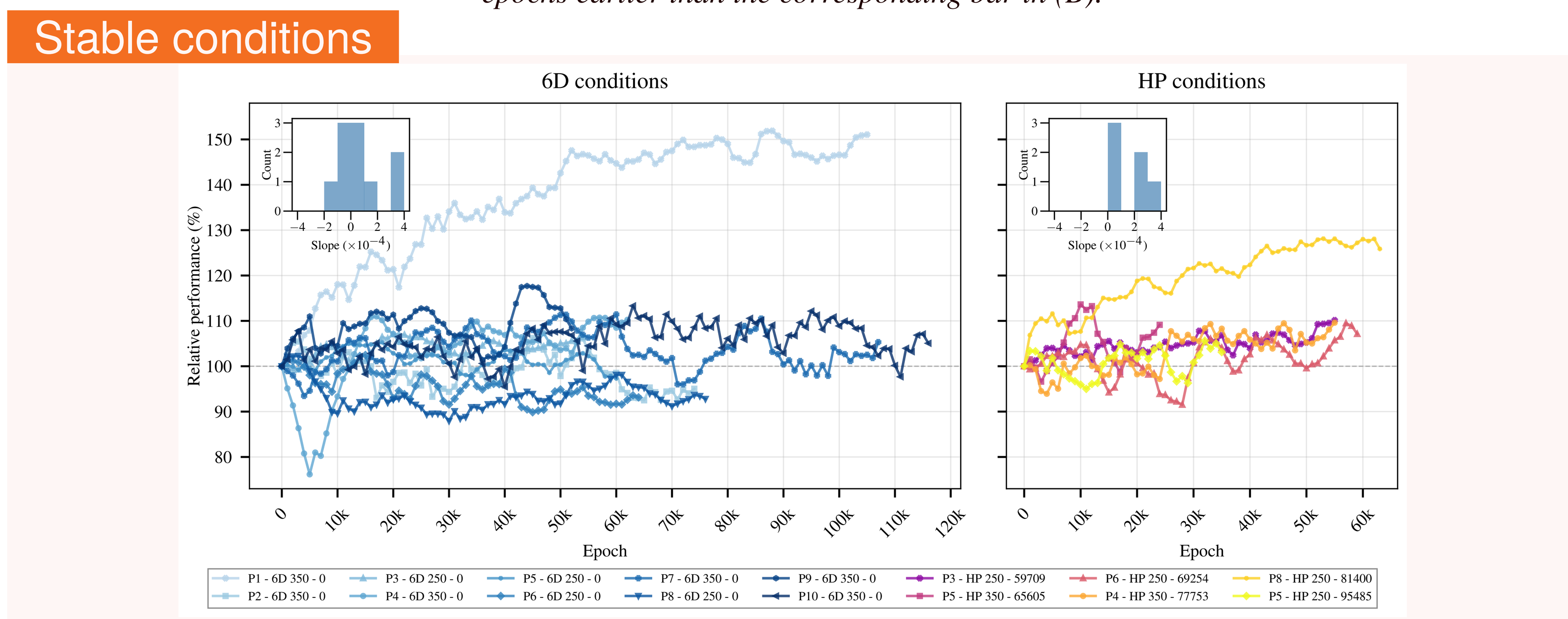
ROC-AUC 5-fold cross-validation performance per patient and over sessions. Gray bars indicate the number of epochs recorded per session. Condition codes describe the combination of spatial information and SOA. HP conditions (red) are more difficult than 6D conditions (blue).

Moving window-based analysis



Performance development over windows of a fixed size. (A) Vertical bars indicate the session boundaries. (B) Overview of experimental conditions used for each patient across all epochs. '6D' refers to the stereo setup with 6 speakers surrounding the patient, and 'HP' refers to the headphones condition. The subsequent number indicates the SOA in milliseconds. Please note that due to the data windowing in the training and validation process, a curve in (A) ends between 5000 and 5999 epochs earlier than the corresponding bar in (B).

Stable conditions



Each line represents a segment, that shares the same experimental condition and contains data of consecutive sessions. Multiple lines (segments) are shown per patient, if multiple conditions have been used during his/her training. The segment legend denotes the patient number, spatial- and SOA condition, and the starting epoch. Left: 6D segments. Right: HP segments. Note that no patient started the online training with the HP condition. The curves depict the performance relative to the first window of each segment. For both conditions, histograms show the distribution of linear slopes for the segments.

DISCUSSION

To our knowledge, **this ERP patient data is unique** in both its size and in that the task difficulty was continuously adjusted to remain challenging throughout 30 hours of training. Not unexpected due to different lesions and language deficits (see [1]), our learning curves revealed **substantial inter-individual differences**. We are aware that **performance improvements may not necessarily reflect language improvements**.

Performance improvements

We observed variable results also w.r.t. the learning curves: Patient P1 serves as **proof of concept, that an improvement of ERP decoding in this patient group is feasible**, while P10 is an example of a stable performance across 30 hours.

Overall, performance increases were observed for all HP segments, but only in six out of ten 6D segments. **Increasing task difficulty may stimulate the learning for patients**, while easier conditions may lack learning pressure or motivation to further improve the ERP classification, as exemplified by P8.

As a task may be more difficult for a patient compared to a healthy participant, **patients may generally start at a lower performance, but can increase it in later sessions**. The few reported multi-session ERP studies with healthy participants support this interpretation: Klobassa and colleagues[3] show good performances from the start for an audio-visual ERP spelling protocol with little to no learning over sessions, while meaningful improvement is observed for a harder auditory protocol. Early improvement in the first three sessions, followed by stable–possibly ceiling–performance, has also been reported for an auditory ERP speller by Baykara et al.[4].

REFERENCES

[1] Musso M et al. Brain communications. 2022;4(1):fcac008.

[2] Vansteensel MJ et al. New England J of Medicine. 2024;391(7):619–626.

[3] Klobassa DS et al. Clin. neurophysiology. 2009;120(7):1252–1261

[4] Baykara E et al. Clin. Neurophysiology. 2016;127(1):379–387.

[5] Schreuder M, Rost T, Tangermann M. Frontiers in neuroscience. 2011;5:112.

[6] Sosulski J, Tangermann M. J of neural eng. 2022;19(6):066001. doi:10.1088/1741-2552/ac9c98

E-mail: michael.tangermann@donders.ru.nl
Lab: https://neurotechlab.socsci.ru.nl/

Poster PDF:

Radboud University Radboudumc

Presented at the 10th Graz Brain-Computer Interface Conference 2026 in Graz, Austria on Wednesday 9th September 2026