Comparative study of training intensity in neurofeedback

Inês Esteves *1, Wenya Nan2, Cristiana Alves1, Alexandre Calapez1, Fernando Melício1,3, and Agostinho Rosa1,4

1Evolutionary Systems and Biomedical Engineering Lab, Institute for Systems and Robotics, Lisbon, Portugal
2Department of Psychology, Shanghai Normal University, Shanghai, China
3ISEL – Instituto Superior de Engenharia de Lisboa, Instituto Politécnico de Lisboa, Lisbon, Portugal
4Department of Bioengineering, Instituto Superior Técnico, University of Lisbon, Lisbon, Portugal

Abstract

Neurofeedback has proved to be useful in many instances. This technique is often used to address both medical issues and performance improvement. Despite the wide range of applications, no consensus has been reached about the optimal training schedule.

In this work, a practical experiment was conducted aiming to compare the effects of intensive and sparse training modalities of enhancement of the individual upper alpha band at Fz, for working memory improvement, in 19 healthy subjects. The INTENSIVE group was submitted to 4 sessions of 37.5 minutes each, during consecutive days, while the SPARSE group performed 15 sessions of 25 minutes along approximately 2 months.

The intensive modality proved to be significantly more effective in increasing the target frequency at Fz within session. However, no significant differences were found across sessions neither regarding cognitive improvements. The distinction between learners and non-learners led to significant results for the two groups across sessions, but only the learners of the INTENSIVE group show significant changes within session. Besides, for this subsample the effects in the INTENSIVE group are significantly stronger both within and across sessions.

*Correspondence to: Inês Esteves, Instituto Superior Técnico, Avenida Rovisco Pais 1, 1049-001 Lisboa, Portugal; inês.esteves@tecnico.ulisboa.pt

This is a preprint version.
The results suggest that longer sessions with shorter trials and more breaks could be advantageous when training healthy subjects. Although further investigation is needed to establish if one modality outperforms the other, better outcomes might be obtained if each session consists in many short trials (in this case 30 seconds) and is performed on consecutive days.

Keywords:
- Neurofeedback
- Training intensity
- Individual upper alpha band
- Working memory
- EEG

1 Introduction

Neurofeedback (NF) relies on the voluntary modulation of brain activity: brain signals are extracted and evaluated in real time, and presented back to the individual in the form of an auditory or visual feedback [Kober et al., 2017]. Since it does not depend on an external electrical or magnetic stimulus, it is expected to act without causing brain dependency [Niv, 2013]. Therefore, it reveals a great potential, as a complementary or alternative therapy, to deal with medical conditions when the conventional treatments are not successful or induce negative side effects [MTA Cooperative Group, 2004; Arnold et al., 2013]. For instance, NF has been used to target attention deficit hyperactivity disorder [Meisel et al., 2013; Mayer et al., 2016; Strehl et al., 2017], which is widely treated with stimulant medication [Strehl et al., 2006b] that may cause problems such as growth suppression, impaired sleep and irritability [National Collaborating Centre for Mental Health (London), 2009; MTA Cooperative Group, 2004]. Another example is the current pharmacological treatment for schizophrenia which is considered insufficient and may cause the limitation and deterioration of functions when used for a long period [Nan et al., 2017]. As a consequence, new treatment options are being pursued, and NF has already shown positive behavioural outcomes when applied to this condition [Bolea, 2010; Nan et al., 2017].

In addition to clinical applications, in recent years, research has spread to non-medical fields and newer protocols have been applied [Gruzelié, 2014; Mirifar et al., 2017]. However, despite the increasing widespread use of this technique, some aspects of NF methodology, such as the number of frequency bands, feedback modality, number of electrodes and training intensity are not standardized [Rogala et al., 2016]. Training intensity comprehends aspects such as the number of sessions, the duration of the subdivisions within one session and the spread of the training over time. For the same application, very different training intensity
parameters have been used. For instance, for schizophrenia, one protocol consisting in a total of 13.5 hours within four consecutive days (Nan et al., 2017) and also another one composed of a total of 130 sessions of 1 hour twice per week, during 18 months, (Bolea, 2010) revealed positive results at the electrophysiological and behavioural levels. Although this example needs an adequate context since it refers to a complex pathology, it illustrates that in some cases much time and resources could be saved, as results could probably be achieved in a more efficient way, increasing also the motivation of the subjects.

Furthermore, several authors attribute poor or unexpected results to insufficient training time (Dias and van Deusen, 2011; Lecomte and Juhel, 2011; Nan et al., 2012). Also, when making a distinction between learners and non-learners, it might even be the case that some individuals considered as non-learners, i.e. that do not respond to the protocol as expected, simply needed more sessions to consolidate results (Wang et al., 2016). Therefore, a deeper knowledge of training intensity would also allow for a more reliable evaluation of the results, by controlling the impact of choices regarding this matter.

In this work, an experiment was performed with healthy subjects to test if contrasting training modalities (distinct with respect to the number of sessions, frequency and duration of each session) produce significant differences. These are quantified regarding electrophysiological changes at the cortical level and performance during a target task. The training aims to increase the individual upper alpha amplitude and, consequently, improve working memory. As there is no optimal protocol regarding intensity, this investigates if the learning process may occur within a short period. This is expected to be more easily accepted by participants as it is lighter and takes less effort regarding scheduling.

To our knowledge, this is the first study that directly compares the effects of two distinct NF training intensity modalities. Although the option of studying several variables (number of sessions, frequency and duration of each session) at the same time may prevent the isolation of a single effect, it allows a preliminary investigation of the impact of distinct intensity factors. The examination of only one intensity parameter would rely on the assumption of a fixed model, dependent on a specific variable, and would reduce the chances of finding any effects, as there are many possible sources of impact on training. Accordingly, we selected two protocols as distinct as possible which would still be feasible in terms of total duration and availability of the participants.
2 Methods

2.1 Participants

A total of 19 healthy subjects participated in this study. Participants were allocated to two different groups: INTENSIVE (n = 9) and SPARSE (n = 10). It was not possible to randomly assign participants to their group: since the training load differed significantly between groups, time constraints had to be taken into account and thus the choice was made according to the participants’ requirements. There was one drop-out in the SPARSE group after 4 sessions, due to incompatibilities with personal schedule. Considering the final sample, the INTENSIVE group consisted in 2 males and 7 females (age: 23.44 ± 2.41 range: 22-30) while the SPARSE group consisted in 6 males and 3 females (age: 27.67 ± 9.81, range: 22-46). The Wilcoxon rank sum test showed no significant difference between groups regarding age (U = 38.5, p = 0.857).

This study was carried out in accordance with the recommendations and guidelines of the Ethics Committee of the CHLN and CAML with written informed consent from all subjects. All subjects gave written informed consent in accordance with the Declaration of Helsinki. The protocol was approved by the Ethics Committee of the CHLN and CAML. Participants were all volunteers and no monetary reward was given for their cooperation with this study.

2.2 Design

The differences between INTENSIVE and SPARSE groups rely only on training intensity. INTENSIVE group performed 4 sessions, of approximately 2 hours each, in 4 consecutive days while SPARSE group performed 15 sessions, around 1h30 each, spread along approximately 8 weeks, with 2-4 sessions per week. A follow-up assessment was carried approximately one month after they completed the training. Both groups were submitted to NF training (upregulation of the UA band with feedback), transfer trials (upregulation of upper alpha (UA) without feedback) and cognitive tests (tests to assess working memory performance). Resting baselines were preceded by a 2-minute relaxed state and assessed both Eyes Open (EO) and Eyes Closed (EC) conditions, with recording of two alternating epochs of 1 minute for each, both at the beginning and at the end of each session. A questionnaire to assess general health state was answered in the first session (before training) and in the follow-up assessment. In every session, participants were also asked to fill a questionnaire in which they rated several parameters referring to their mental state during the session.

Diagrams illustrating the training workflow for the INTENSIVE and SPARSE groups are represented in Figure 1 top and bottom, respectively. Although the cognitive tests were
performed at the beginning or at the end of one session, they are represented outside the respective session. This is done to facilitate the understanding of the schematics, since these evaluations were not part of every session.

Participants in the INTENSIVE group underwent sessions with 5 sets of 3 blocks each. The blocks were composed by 5 trials of 30 seconds, with 5-seconds intervals between them. Transfer trials were introduced from session 2 to 4, during which no feedback was provided. A transfer trial for adaptation, with a duration of 30 seconds, was introduced at the end of the 3rd set of blocks in the 2nd session. The following transfer periods consisted in two trials of 30 seconds and were performed after the 5th set of the 2nd session, after the 3rd and 4th sets of the 3rd session and after the 3rd and 4th sets of the 4th session. The total NF time in each session was 37.5 minutes (not taking into account transfer trials and intervals between trials). Cognitive tests were performed after the pre-training baselines in the first session and after the post-training baselines in the last session. The follow-up assessment was similar to the first training session, with the exception that the training period consisted in a single training block, followed by a 30-second transfer trial. Figure 2 (Top) illustrates the plan for each session of the INTENSIVE group. Again, in order to facilitate the representation, intermittent assessments are not included in the schematic, which in this case correspond to the transfer trials, whose temporal organization within training varies among sessions.

Participants in the SPARSE group underwent sessions with a single set of 5 blocks. Each block consisted in 5 trials of 60 seconds each, with 5 seconds of interval between them. Transfer trials were introduced only in the 11th session and lasted until the last session. In every session, after the 5th NF training block, a transfer block was performed, with 2 transfer trials of 1 minute separated by 5 seconds of interval. Cognitive tests were performed after the pre-training baselines in the first session and after the post-training baselines in the 5th, 10th and 15th sessions. As in the INTENSIVE group, the follow-up assessment also resembled the first session, differing only in the existence of a single training block which was followed by a 1-minute transfer trial. Figure 2 (Bottom) illustrates the plan for each session of the SPARSE group, transfer trials are also not included.

2.3 Signal acquisition

The acquisitions were carried out using Somnium software (Rodrigues et al., 2010), in a room provided by the Evolutionary Systems and Biomedical Engineering Lab (LaSEEB), a research lab of Institute for Systems and Robotics (ISR), at Instituto Superior Técnico (IST), University of Lisbon. Electrodes were placed according to the International 10-20 System, using the left and right mastoids as references for common mode rejection and the
middle of the forehead as ground. Relevant signal was recorded, with a sampling frequency of 250 Hz, from 20 electrodes: Fz, Fp1, F7, F3, T3, C3, T5, P3, O1, Cz, Pz, Oz, Fp2, F8, F4, T4, C4, T6, P4 and O2, and amplified by the EEG amplifier Vertex 823 (produced by Meditron Electromedicina Ltda, São Paulo, Brazil), with an analogical band-pass filter between 0.1 and 70 Hz. The impedance was measured during the process, guaranteeing that all electrodes had an impedance below 10 kΩ. Subjects were asked to remain as still as possible and also to avoid excessive blinking and abrupt movements.

2.4 Protocol

Although the main focus of the present work is the study of training intensity effects, it also tries to evaluate the outcomes of distinct training intensity modalities regarding the enhancement of working memory (so that healthy participants felt more engaged with the tasks). For that purpose, training consisted in the enhancement of the individual upper alpha band at Fz, as both the upper alpha band and the frontal area are associated with memory functions (Zola-Morgan and Squire, 1993). The evidence of a link between increased upper alpha band activity and good working memory performance (Klimesch et al., 1999; Escolano et al., 2011; Hsueh et al., 2016) lead to the decision of training individuals to increase the activity in this band. The use of the Individual Alpha Band (IAB) was based on Klimesch (1999) that invokes a large variability of the boundaries of this band among subjects to justify the need for a personalized definition.

2.4.1 Individual upper alpha measurement

The first baseline measurements were used to define the IAB of each individual. The power spectrum density is estimated using the Welch’s method (Welch, 1967), with an overlap of 10% and a segment length of 5 seconds, in Somnium. The crossings between EO and EC spectra provide the frequency boundaries Lower Transition Frequency (LTF) and Higher Transition Frequency (HTF). The UA band is defined as the frequency range between the individual peak alpha frequency, which was considered to be equivalent to Individual Alpha Frequency (IAF), and the HTF.

2.4.2 Neurofeedback training

Feedback

The EEG training platform integrated in the Somnium software was used to perform NF training. The feedback is established according to the achievement of two goals and makes use of two three-dimensional objects against a grey background, a sphere and a cube, that
change according to previously defined settings and react to subject’s EEG in real-time (Rodrigues et al., 2010). Goal 1 depends on the relation between the EEG measurement that is being evaluated and a pre-defined threshold value. In this case, since the interest is to enhance the upper alpha band, it is considered that Goal 1 is achieved every time that the relative amplitude of this band is above the threshold. This makes the sphere change from white to purple and increase in size and number of sides. Goal 2 is achieved everytime that Goal 1 is sustained for more than 2 seconds, making the cube go up until Goal 2 is no longer reached. Using $X(k)$ to denote the frequency amplitude spectrum computed using the Fast Fourier Transform with a sliding window of 2 seconds and shifts every 0.125 seconds, the relative amplitude as presented in Wan et al. (2014) is given by:

$$\text{Relative Amplitude} = \frac{\text{Band Amplitude}}{\text{EEG Amplitude}}$$

In this case, for the upper alpha relative amplitude:

$$UA \text{ Relative Amplitude} = \frac{\sum_{k=IAF/\Delta f}^{HTF/\Delta f} X(k)}{\sum_{k=3.99/\Delta f}^{30/\Delta f} X(k)}$$

(2)

The frequency resolution is represented by $\Delta f$. The EEG amplitude is measured considering the frequencies between 3.99 and 30 Hz (lower frequencies were discarded to mitigate the effects of blinking). The baseline amplitude values were defined as described in the previous subsection and maintained along the sessions for both groups, as we assume there should not exist significant intra-variability.

All subjects started with a threshold value of 1, which was adjusted according to individual performance evaluated by the total amount of time during which Goal 1 was reached. If the percentage of time during which the feedback parameter was above threshold exceeded 60%, the threshold was increased by 0.1. If this percentage was lower than 20%, the threshold was decreased by 0.1. This was done in order to keep it challenging if the performance is considered good and, if the opposite happened, to allow the subject to find the most successful mental strategies without losing motivation along the process. The aim was to maximize the effect of the feedback.

**Training**

Participants were informed about the existence of the two goals during NF training and told to focus primarily on the sphere, trying to make it purple, and as round and big as possible. They should use a single mental strategy per block, so that the effects of that strategy
could be isolated in order to rate its effectiveness. During the first blocks, participants were encouraged to try different strategies in order to understand which ones produced better results and then repeat them afterwards. They were allowed to make short breaks between blocks, if needed, while the practitioner defined the parameters for the following block and wrote down the mental strategy that was used.

In the INTENSIVE group, between the sets of blocks there was a larger break that also helped to check the average time spent above threshold and reset the NF threshold when necessary. In the SPARSE group, threshold resetting only occurred at the end of each session.

Transfer trials were performed to assess the ability to control cortical activity without feedback (Rockstroh et al., 1993; Siniatchkin et al., 2000).

### Mental strategies

Although the participants were not encouraged to use any specific strategies, some examples were provided when they asked for them. Since they were allowed to choose the more suitable strategies for them, the preferentially applied strategies varied a lot. The most successful strategy of each session (to which corresponded the block with highest UA amplitude) was collected for each individual. After gathering all the best strategies for every subject, 49 distinct strategies were found which were then grouped into six categories: “feedback”, related with feedback display and the screen; “imagination”, for fantasizing with unreal episodes; “memories”, for recalling past experiences; “mental”, when performing tasks that involved mental effort; “motor”, for thinking about performing physical activities; “relaxation”, attempting to relax the body and mind (for example, with breathing exercises). For both groups, “relaxation” was the preferred category (INTENSIVE: 38.89% and SPARSE: 34.07%). When counting the total number of different best strategies there was not a distinction between repeated use by the same participant or single use by several participants.

### 2.5 Measures

#### 2.5.1 Working-memory performance

So that the effects of NF training in working memory performance could be analysed, allowing also a comparison between the two training intensity modalities, different cognitive tests were carried out: Digit Span (Forward and Backward) and N-Back (Numbers and Letters). During the Digit Span test (To et al., 2016) participants had to recall several sequences, with increasing length in each trial, from 3 to 13 digits. Subjects were asked to introduce the digits in the order by which they appeared (forward digit span) and by the reverse order (reverse digit span). When the test finished, two performance measures were provided: digit
span, the maximum amount of digits of a single sequence that one was able to remember, and the score, the percentage of correct correspondences between the real sequence and the one introduced by the subject.

During N-Back test (Kirchner, 1958) the individuals were required to monitor a series of numbers/letters and identify if the current one was the same that was presented n trials before, with n = 3. Twenty-three trials were performed, which resulted in 20 answers. Each digit was shown for a maximum of 3 seconds, during which the participant could answer, and there was an interval of 2 seconds between trials.

2.5.2 Questionnaires

In order to keep record of how factors such as concentration, motivation, sleepiness and stress affect training, a questionnaire to assess these factors was performed. For this purpose, a rating scale was used to evaluate the frequency of the four mentioned states/sensations during training: 1 - never, 2 - rarely, 3 - sometimes, 4 - frequently, 5 - always.

Furthermore, the 36-Item Short Form Survey (SF-36) questionnaire, which has been validated for the Portuguese population, was performed to assess different domains of health state and quality of life (Ferreira et al., 2012; Corporation, 2017). It was performed in order to ensure that the participants did not have abnormal health conditions and at the same time compare their health state before and after the NF training.

2.6 Data pre-processing and extraction

The pre-processing and extraction of information from the raw EEG was performed using Somnium tools. It was necessary to remove some recorded periods which contained artifacts that would mislead the analysis. With the exception of eye movements, the artifact removal was performed manually, for all baseline measurements of each participant. Individual reports were transferred from Somnium to Excel files and extracted with MATLAB software (version R2015b). The studied relative amplitudes correspond to the frequency bands presented in Table 1.

2.7 Data treatment

The treatment of the data for this work, including the visualization of plots as well as statistical analysis, was performed using the MATLAB software. The topographic distributions were generated using Fieldtrip (Oostenveld et al., 2011), an open-source MATLAB toolbox.
2.7.1 Evaluation of training performance

Since the participants had different baseline amplitude values and the electrode impedance varied from session to session, as the only restriction imposed was that values were under 10 kΩ, relative amplitude values for each individual session were normalized. The normalization was enforced through division by the corresponding pre-training baseline with EO. Therefore, for simplification, normalized relative amplitudes will be hereafter referred as amplitudes only.

There is not a standard approach for the evaluation of training performance. Several authors relied on clusters of sessions or representative sessions, such as the first and the last, to extract conclusions about training effectiveness (Strehl et al., 2006a; Leins et al., 2007; Staufenbiel et al., 2014; Hsueh et al., 2016). On the other hand, others use measures that consider the progress along the whole training (Escolano et al., 2011; Nan et al., 2012). In this case, in order to evaluate the evolution of the amplitude of frequency bands along time, both within session and across sessions, distinct indexes were used to capture different aspects. We derived two types of measures, one that explores the overall trend and another to take into account the small variations that may occur. For $S$ sessions and $B$ blocks, denoting the $i$−th session by $s_i$ and the $j$−th block of that session by $b_{j,i}$:

- **Within Session**
  - $W1$: this measure is based on Wan et al. (2014) and aims to quantify the mean change within session across all training sessions:

    $$ W1 = \frac{\sum_{i=1}^{S} \sum_{j=2}^{B} (b_{j,i} - b_{1,i})}{S \times (B - 1)} $$  \hspace{1cm} (3)

  - $W2$: is the slope of the linear regression that describes the evolution of the amplitude of a certain frequency band along blocks, averaged across sessions. Considering $y = m_i x + b_i$ as the trendline for $s_i$:

    $$ W2 = m_{\text{average}} = \frac{\sum_{i=1}^{S} m_i}{S} $$  \hspace{1cm} (4)

    in which $y$ corresponds to the amplitude, $m_i$ is the slope, $x$ is the block number and $b_i$ is the y-intercept, which will depend mostly on each subject’s characteristics.

- **Across Sessions**
  - $A1$: the amplitude change of the last two sessions relative to the first two sessions.
\[ A1 = \left( \frac{s_{S-1} + s_S}{s_1 + s_2} \right) \] (5)

- \( A2 \): corresponds to the slope of the linear trendline computed for the evolution of frequency amplitude along sessions, therefore taking into account the variations that may occur in between. Taking \( y = mx + b \), \( A2 \) corresponds to \( m \)

\[ A2 = m \] (6)

with \( y \) corresponding to the relative amplitude of a certain frequency range, \( x \) representing the session number and \( b \) standing for the y-intercept, which, as in the previous case, will rely on each subjects’ intrinsic characteristics.

The distinction between learners and non-learners relies only on the training performance for the target frequency band, in this case the UA band. Similarly to what was mentioned for the evaluation of training performance, there is also not a consensus regarding how to distinguish learners from non-learners and distinct methods have been applied (Escolano et al., 2011; Hsueh et al., 2016; Mayer et al., 2016). In the present work, \( A2 \) measure being positive (positive slope) was chosen as the criteria to classify a participant as learner, as it reflects the learning trend across time and thus summarizes the performance. Five learners were identified in the INTENSIVE group (56%) and six in the SPARSE group (67%).

2.7.2 Evaluation of cognitive effects

The results of Digit Span Test and the N-Back test were used to evaluate the effects on cognition, at the working-memory level. For the Digit Span Test, two indicators were chosen “Span” and “Score”. Regarding the N-Back Test, the performance was evaluated based on the “Number of Successes”. The results obtained at \( t_i \) (initial session), \( t_f \) (final session) and \( t_{fu} \) (follow-up session) were used to evaluate the effects along time and examine if the training influenced them.

2.8 Statistical analysis

This experiment was performed with a small sample (total \( N = 18 \) participants in the final sample), and several results are analysed individually for each group (\( N = 9 \) for each group). Shapiro-Wilk test (Shapiro and Wilk, 1965) was used in order to assess if variables under examination came from a normally distributed population. Since in several cases normality was not verified, an option was made to rely only on non-parametric tests.
To analyse the median of a group or the difference between two time points within one group, Wilcoxon signed rank test was used (Conover and Iman, 1981). Right-tailed tests were employed when there was a prior hypothesis of an increase, as was the case for the UA amplitude within and across sessions as well as (post−pre) baseline measurements. For validating of the use of the individual upper alpha band, the boundaries IAF and HTF were not compared with zero, but with standard values 10 and 12 Hz, respectively (Klimesch et al., 2005).

In order to make comparisons between groups, the Wilcoxon rank sum test was used, mostly to test if the medians considering all the participants of each group differed significantly (Chernick and Friis, 2003). A significance level of 5% was considered.

In the case of topographical analysis regarding the UA amplitude, since multiple comparisons were performed as there were 20 electrodes, the false discovery rate (FDR) was controlled with the Benjamini-Hochberg procedure (Benjamini and Hochberg, 1995). Considering \( m \) hypotheses \( H_1, H_2, ..., H_m \) to which correspond the p-values \( P_1, P_2, ..., P_m \), the p-values are sorted in ascending order \( P^{(1)} \leq P^{(2)} \leq ... \leq P^{(m)} \), with \( P^{(i)} \) corresponding to \( H^{(i)} \). For a given threshold \( q \), one should find \( k \), the largest \( i \) for which:

\[
P^{(i)} \leq \frac{i}{m} q
\]

and reject all the hypotheses \( H^{(i)} \) with \( i = 1, ..., k \). The threshold \( q \) was defined as 0.05.

3 Results

3.1 Individual alpha band

Concerning only the initial IAB, at \( t_i \), the results indicate that the groups did not differ significantly regarding LTF (\( U = 39, p = 0.914 \)) and IAF (\( U = 30, p = 0.372 \)). However, the difference for HTF values was considered significant (\( U = 17.5, p = 0.041 \)), at \( t_i \).

In order to assess the effect of training on the intra-variability of the individual alpha band, it was tested if the median of the change (\( t_f - t_i \)) was different from zero, for each group individually. There was only evidence of a statistically significant difference between HTF measured at \( t_i \) and \( t_f \) in the SPARSE group (\( W = 33, p = 0.039 \)). However, it is possible that this change was not caused by the training intensity modality itself, but rather by intrinsic small changes along time. These were more likely to appear in this group since the \( t_i \) and \( t_f \) are separated by approximately 2 months. At \( t_f \), there were no significant differences between groups for any of the frequencies.

Although a common fixed reference range for the upper alpha band is from 10 to 12 Hz
(Klimesch et al., 2005), this experiment used the individual upper alpha band, so that each subject had a personalized intervention. To examine if the individual range was similar to the reference one, it was tested if the IAF and HTF of both groups (as a whole) differed significantly from 10 and 12 Hz, respectively. At \( t_i \), IAF was not considered significantly different from 10 Hz (\( W = 128, \ p = 0.064 \)). While for HTF the difference from 12 Hz was considered significant (\( W = 140, \ p = 0.003 \)). Group values at \( t_i \) were LTF: \( Mdn = 8.25, \ IQR = 7.9 - 8.8 \), IAF: \( Mdn = 10.4, \ IQR = 9.7 - 11.2 \) and HTF: \( Mdn = 12.9, \ IQR = 12.5 - 13.3 \).

3.2 Training

The analysis of NF training results focuses on the changes that occur both along the time course of the training experiment and within each session. The evolution of the relative amplitude of UA was examined to assess trainability while changes regarding other frequency bands are shown so that independence may be studied, as suggested by Zoefel et al. (2011). The topographical effects of training were also inspected in order to understand whether they are restrained to the training location, Fz, or if there was a spread to other areas.

3.2.1 Across sessions

Training location

The evolution of the UA amplitude across sessions for both groups is depicted on Figure 3, with different scales, as the number of points is much smaller for the INTENSIVE group. For both groups there are amplitude fluctuations along time for the whole training period, although the difference between the initial and final values is larger for the SPARSE group. Accordingly, in Figure 3, the A1 measure shows a higher median for the SPARSE group and its distribution is shifted upwards when compared to the INTENSIVE group. For the A2 measure, both medians are close to zero and the distribution is highly concentrated for the SPARSE group, which indicates that this measure might have been a poor choice to discriminate between learners and non-learners since the differences are minimal. Overall, no significant effects across sessions were found for both A1 and A2, either within groups (Table 2) or between them (Table 3).

Topographic distribution

Concerning the measures of performance across sessions, represented in Figure 4, the topographical results are according to what was previously noted from Figure 3 as there is an area of increased A1 around Fz for the SPARSE group. However, there are values of similar
magnitude for the INTENSIVE group in the right frontotemporal and left temporal areas, which were not expected due to training. This pattern is also visible for the A2 measure, for which the values in the case of the SPARSE group are around zero. Nonetheless, the only significant difference between groups is regarding A2 at electrode F8, in the frontal area.

3.2.2 Within session

Training location
Despite the fact that the within-session performance might change along sessions, we still analyzed them globally for all sessions due to their similar progressions over blocks. Therefore, Figure 5 refers to one representative session, in which the points represent the mean value for a certain block across sessions, considering the median for all subjects. While there is no visible trend for the SPARSE group, the amplitude tends to increase along blocks for the INTENSIVE group, although it slightly decreases in the last two blocks.

The distribution of the learning measures may be visualized in Figure 5. Measures regarding changes within sessions are not significantly different from zero for any of the frequency bands of the SPARSE group, as presented in Table 4. The medians of both W1 and W2 for UA were considered significantly larger than zero for the INTENSIVE group, which may be observed both in this figure and in Table 4. Moreover this group presents significant differences for both measures when compared with SPARSE group ($U = 68, p = 0.014$ and $U = 72, p = 0.004$, respectively), visible in Table 5. However, for the INTENSIVE group, W1 and W2 are also significantly different from zero for LA1, IAB and Beta2, and W2 only for LA2. These changes within the INTENSIVE group are also visible when comparing both groups, with significant differences regarding LA1, IAB and Beta2. Therefore, although only the INTENSIVE group reveals changes within session, these are not restricted to the target band.

Topographic distribution
Comparing the topographic distributions, in Figure 6 it is visible that, for both measures, the values are higher for the INTENSIVE group, specially in the occipital, temporal and parietal areas, with significant differences between groups. Although not directly targeted by the training, an increase of UA amplitude in these areas might be more easily promoted as the alpha waves occur predominantly in the occipital area. Nevertheless, as expected due to training, the significant differences between groups occur mainly in the frontal area.
3.2.3 Learners

When evaluating only the learners, both A1 and A2 for the UA band have much larger medians than non-learners for both groups, as it was expected considering the rationale used to make this distinction. UA amplitude showed significantly increased performance across sessions in both INTENSIVE (A1: $W = 15$, $p = 0.031$; A2: $W = 15$, $p = 0.031$) and SPARSE groups (A1: $W = 21$, $p = 0.016$; A2: $W = 21$, $p = 0.016$). Between groups, A2 regarding UA differs significantly ($U = 28$, $p = 0.017$), with the median being larger in the case of the INTENSIVE group.

Within sessions, only W2 for UA amplitude in the INTENSIVE group is significantly larger than zero ($W = 15$). This value is significantly different between groups ($U = 30$, $p = 0.040$). Hence, even when considering only the subjects with a positive trend along time, the effects within sessions are exclusive to the INTENSIVE group and along sessions this group reveals stronger changes.

3.2.4 Transfer trials

In order to assess if there were a relationship between training performance and the ability to achieve the goal without feedback, the correlation between the A2 measure of training performance during neurofeedback training and transfer trials was computed. For each participant, only the best transfer trial was chosen, which consisted in the one for which the UA amplitude was the largest. Using the Pearson’s correlation coefficient, no significant correlations were found (INTENSIVE: $r = 0.278$, $p = 0.469$, SPARSE: $r = -0.103$, $p = 0.793$). No significant differences were found between groups regarding the best trials ($U = 49$, $p = 0.475$). Thus, it seems that the participants’ ability to control their activity outside the training environment was independent of the performance during training.

3.2.5 Follow-up

Three subjects had to be excluded from follow-up, two due to lack of availability and one due to technical issues during the follow-up session. Thus, the follow-up assessment was performed only on 15 subjects, 8 of the INTENSIVE group and 7 of the SPARSE group. Concerning the UA relative amplitude, regarding the NF period, for both groups most of the participants had lower values during follow-up than during the last session. This is more pronounced in the INTENSIVE group, for which the median is lower and the dispersion to lower negative values is larger. However, the differences between the two time points were not significantly different from zero for any of the groups (INTENSIVE: $W = 6$, $p = 0.109$ and SPARSE: $W = 19$, $p = 0.469$). The results are also not considered significantly different.
between groups ($U = 2$, $p = 0.054$), yet the obtained p-value is close to 0.05, therefore this outcome should be taken into account.

### 3.3 Baselines

In the data discussed in the present subsection and shown in Figure [Figure 7], the NF amplitudes were not normalized, to allow a comparison between the values obtained during training and baselines. While for the INTENSIVE group the median for EO post session was larger than before the session for 3 of the 4 sessions, there is not a marked tendency for the SPARSE group. Moreover, there is no clear trend along time either during EO or EC, and for any of the groups.

The UA amplitude is predominantly larger during NF than during EO baselines, both pre and post NF. Nonetheless, the values rarely surpass the ones obtained during EC. When evaluating the differences within session, i.e. post-baseline $-$ pre-baseline, no p-values under 0.05 were obtained for any participants, both for EO and EC in the two groups. Therefore there is not enough evidence to conclude about significant effects of training on the baselines within session.

The slope of the trendline across sessions was analysed to determine if there was a positive evolution and also to check differences between groups. No p-values under 0.05 were obtained both for EO and EC, either for the within group or the between groups analysis.

The correlation between pre-baseline values of the first session, prior to any neurofeedback training, and A2 measure was computed to test if this could be used as a predictor of training performance. No significant correlation was found for EO ($r = 0.157$, $p = 0.534$) or EC ($r = 0.105$, $p = 0.678$).

### 3.4 Cognitive tests

Medians of change ($t_f - t_i$) for both groups were not considered significantly different from each other for any of the cognitive tests. However, within group there were score changes whose median was considered significantly larger than zero, namely regarding the Digit Span for both groups and the N-Back with letters only for the SPARSE group.

The correlation between the performance in cognitive tests and training performance was assessed in order to understand if the results could be seen as NF training effects. Although the subjects in the SPARSE had more chances to practice, the period between assessments was also longer, and there was no evidence that this favoured their performance. A significant correlation was found between the score change in Digit Span Test (Reverse) and A2 for the INTENSIVE group ($r = 0.740$, $p = 0.023$).
Respecting the follow-up, regardless these effects on the EEG, while for the INTENSIVE group the median score and span of the Digit Span (Forward) test increased in the follow-up session, they decreased for the SPARSE group. For the N-Back test, the performance level either slightly increased or remained the same, and the same happened for the Digit Span (Backward) test with exception of the Span for the INTENSIVE group, which decreased.

4 Discussion

4.1 Individual alpha band

The individual alpha band was used in this experiment instead of a fixed range of values. This decision allows for a more personalized approach and thus should provide a more appropriate training. Yet, it may be also a source of error and variability, as the measurement relied on a single baseline (four interleaved minutes of EO and EC condition) and the manual marks were not straightforward for some of the subjects. Along time, there were no significant changes of the IAF, as in Zoefel et al. (2011) and Escolano et al. (2011). The only case of intra-variability occurred for the HTF of the SPARSE group. Although this could have been an effect caused by training, which should be further studied, the difficulties in manually marking the frequencies might have triggered this variation.

4.2 Trainability

Zoefel et al. (2011) defines trainability as the existence of effects due to training in the target frequency band, which for them was also the upper alpha band. Considering the whole training for all participants, there was no evidence of significant changes in the UA across sessions for any of the groups. Across sessions, none of the training intensity modalities revealed to be effective in producing a growing increase of the UA amplitude.

However, making the distinction between learners and non-learners, and thus considering 56% and 67% of the participants of the INTENSIVE and SPARSE groups, respectively, the results point to trainability for both groups. The differences between groups indicate that the INTENSIVE group achieved a significantly higher increase of UA amplitude than the SPARSE group. Equivalent results across sessions were also obtained by Zoefel et al. (2011) and Escolano et al. (2011) for learners (79% and 60%, respectively). Both performed 5 sessions with 5 training blocks of 5 minutes each in five consecutive days. The training load in each session was equivalent to the one for SPARSE group, but the training was done intensively which, combined with the results for this study for learners, suggests that sessions on consecutive days may be more suitable for UA increase.
Within session, for the INTENSIVE group, W1 and W2 measures were considered significantly larger than zero regarding UA, which means that this group accomplished a good training performance within sessions. The mean for all sessions decreases for the last 2 blocks suggesting that a larger training session would not benefit training. With that being true, taking out one set of blocks could even be advantageous. Nonetheless, it cannot be discarded that the participants could always be more distracted in the last blocks, no matter the total number, as they might be anticipating the end of the session. For the SPARSE group, however, W1 and W2 were not significantly larger than zero nor there were any visible changes within session for the UA band. Yet, another study, also performing 5 trials with 5 minutes each, found a significant positive tendency for UA (using a metric similar to W2), though only one session was performed (Escolano et al., 2012), for what it cannot be equitably compared with the present case (15 sessions).

These differences between groups were significant, with the INTENSIVE group obtaining a better performance within session. The equivalent training time of the one received by the SPARSE group, for a single session, is 10 blocks. Since the increasing trend for the INTENSIVE group starts at the first blocks and differences between groups are already visible if one considers only these 10 blocks, it may be concluded that the disparity is mainly due to block duration and not to the total number of blocks. Therefore it seems that smaller blocks, in this case 30 seconds, might be more satisfactory for training when compared to larger ones, of 60 seconds. Additionally, larger breaks might also be beneficial, alleviating the burden of a longer session. Nevertheless, in the work of Kober et al. (2017) both stroke patients and healthy controls were able to regulate upper alpha activity within session, with each session consisting of 6 feedback runs of 3 minutes. Even when considering only the learners, within session, there are only significant effects of training for the INTENSIVE group and these are considered significantly stronger than in the SPARSE group.

### 4.3 Independence

The ability to alter the UA band independently of other frequency bands was evaluated only regarding NF training. Considering all subjects, there were no significant changes across sessions for other frequency bands for any of the groups. Notwithstanding, analyzing across sessions for learners only, independence was accomplished for the INTENSIVE group, but not for the SPARSE, though the differences are for LA2 (close to UA) and IAB (which includes UA). These results are not in line with what was obtained by Zoefel et al. (2011) and Escolano et al. (2011), for whom independence was accomplished regarding training. However, their conclusions are based in different computations. Zoefel et al. (2011) compared the active
pre-baseline of the first session with the last training block (on the 5th session), testing the
differences regarding only lower alpha (defined as IAF - 3 to IAF - 1 Hz) and lower beta
(defined as IAF + 3 to IAF + 5 Hz), for which no significant effects were found. In the
case of [Escolano et al., 2011], the comparison was made between the pre-active assessment
block of the first and the last session, and the same bands were tested, with no significant
differences for any of them. Furthermore, they performed less sessions which, combined with
the independence in the case of the INTENSIVE group (also with few sessions), may indicate
a larger difficulty in maintaining specificity along time.

Analyzing the results within session, for the SPARSE group there are no significant effects
on other frequencies, either considering the whole group or only the learners. However,
contrary to what was verified along sessions, for the INTENSIVE group there are significant
changes for LA1, LA2 and IAB and also for Beta2. Hence, it is considered that UA was
not trained independently from other frequency bands within session for this group, though
independence was accomplished when taking into account only the learners. It is possible
that the increased LA1 in the INTENSIVE group is due to the need to maintain attention
for a longer period of time, which could also cause the participants to remain stiller along
the session and have lower Beta2 (as this band is more prone to muscle artifacts).

4.4 Baselines

Contrarily to what was obtained by [Escolano et al., 2011] with passive baselines, for the
upper alpha band, the effects of training did not show up on the baselines along time, neither
for EO nor for EC. Nonetheless, they considered only the learners and compared the pre-
block of the first and last sessions while this study analyzed the slope along time. In the
work performed by [Dekker et al., 2014], 15 sessions with 3 blocks of 8 min were performed
on consecutive working days with the aim of increasing the power of the alpha band. The
upper alpha power during a passive baseline showed a significant increase between session 1
and 10, but decreased afterwards, which was not observed in this study.

Considering active baselines, both trainability and independence were observed by [Es-
colano et al., 2011] and by [Zoefel et al., 2011] (where only active baselines were recorded).
Likewise, [Escolano et al., 2012] found significant differences between the NF and the control
group concerning active screenings performed before and after a single session. Yet, no sig-
nificant effects were found within group when comparing the pre-baseline of the training day
and the assessment made on the day after. [Navarro Gil et al., 2018] also found significant
UA power enhancement during task-related activity comparing the initial and final EEG
screenings after 6 sessions (5 trials with 4 minutes each), with 2 sessions per week. Despite
this, no effects were observed regarding the EC resting state.

Regarding clinical applications, Escolano et al. (2014) performed 18 sessions over a month, each with 5 trials of 4 minutes, targeting attention-deficit hyperactivity disorder, and the observed effects were exclusively for task-related activity baseline (EO), they did not show upon resting state (EC). Lavy et al. (2018) completed 10 sessions of 30 minutes (10 trials of 3 minutes each) on mild cognitive impaired subjects in order to increase the individual upper alpha. Although there were changes in the peak alpha frequency and memory scores, there was not a significant change of UA power following training.

These results suggest that, regardless of the intensity modality, training effects may not appear on resting state baselines. In this case, only a passive baseline was recorded and perhaps the introduction of an active baseline would allow for a better evaluation of the effect of the need to pay attention to a task, which comes alongside with NF training. However, since the active tasks visually resemble the NF training (Zoefel et al., 2011; Escolano et al., 2011, 2012, 2014; Navarro Gil et al., 2018) it cannot be discarded that, after the participant has experienced training, visualizing feedback elements may trigger changes in activity even if the training is not being applied, and it is considered that this issue should be further investigated.

No predictive value was obtained from the EO and EC regarding the performance during training for UA enhancement, contrarily to what was expected following the work of Wan et al. (2014). They discovered a significant correlation between resting alpha amplitude and learning indices (greatest correlation with the slope of the logarithmic regression along time of the alpha amplitude during sessions). Notwithstanding, the current work used a linear regression instead of logarithmic and was specific to a sub-range of the alpha band. Besides, the fact that the training schedule is also different may as well contribute to this discrepancy.

4.5 Follow-up

Despite the fact that the results across sessions (for learners only) and within sessions indicate that the intensive training may be more effective, the follow-up outcomes suggest that the learned control may fade more easily. In fact, the (negative) difference between the follow-up and last session is larger for the INTENSIVE group. Since the training is carried out during 4 days only, while for the SPARSE group it lasts for almost 2 months, it is possible the lack of repetition causes the participants to be more prone to forget how to train. For this reason, occasional “boosting” sessions might be useful to maintain the performance level (Angelakis et al., 2007). It also has to be taken into consideration that the follow-up session did not have the same duration as the regular session, and was in fact much shorter. For that reason,
it might not mirror the training and, in this case, since INTENSIVE’s trials were shorter it might not have been enough time for the effects to show.

4.6 **Cognitive tests**

No significant differences were found between groups either for the Digit Span or the N-Back test. However, there were significant effects within each group. Therefore, NF training may have caused improvements but training intensity probably did not influence the results directly. This suggests that the differences regarding NF training were not sufficient to translate into performance. Furthermore, the subjects were young and with high level of education, so there might not be enough space to much greater improvements in one of the groups.

4.7 **Remarks and limitations**

As training intensity encompasses numerous parameters, such as trial duration, trial number and frequency of the sessions, more groups or distinct experiments would be needed in order to test the influence of each of them on training performance. In this work, an effort was made to choose two contrasting protocols and, at the same time, design them in a way that would be suitable for the participants. However, this causes a problem when trying to isolate the effects of a single parameter and understand, for example, whether the INTENSIVE group achieved a better performance within session due to shorter trials or to the larger number of trials. Furthermore, ideally both groups would be subjected to the same total NF training time, but again by virtue of design constraints it was not feasible, as it would either require an overwhelming load for the INTENSIVE group or too few sessions for the SPARSE.

The follow-up was designed as a summary of the training, containing a single training block and one transfer trial, besides the cognitive tests and health questionnaire. This choice was made in order to lighten the session so that more volunteers would be willing to participate in the follow-up. However, it impaired a direct comparison with the previous sessions.

The use of the baselines of each session to compute the normalization may introduce bias in the analysis of NF results if there is a significant effect in baselines along time. Although that was not the case in the present experiment, alternative ways should be explored.

Furthermore, the fact that A2 values are highly concentrated for the whole SPARSE group, and also the contrast when comparing learners and non-learners across or within sessions, point to the conclusion that this might not be the most suitable way of measuring performance. Otherwise, the performance could be assessed by comparing the first trials of
the first session with the last trials of the last session, which considers both the evolution along time and along session.

Though the choice of the training location was related with working memory, it is also suggested that in future studies the training location is changed to more posterior areas. The amplitude of training frequencies in the alpha range may be particularly difficult to increase frontally since naturally they are decreased relative to posterior areas. Actually, UA enhancement has been tested by addressing P3, Pz, P4, O1 and O2 (Escolano et al., 2011; Zoefel et al., 2011; Escolano et al., 2012; Navarro Gil et al., 2018).

Moreover, this experiment was performed with healthy subjects, and a specific band and training location were used. Choosing a clinical target or different protocols may cause significant changes regarding the suitability of a certain intensity modality. Additionally, a control group, as well as a larger sample size are recommended to mitigate the impact of inter-individual variations and allow for strongly supported conclusions.

5 Conclusions

In the present study, an experiment of NF training for the increase of the activity in the UA band was performed, using two different training intensity modalities. Although further investigation is needed to establish if one modality outperforms the other, the results suggest that better outcomes might be obtained if each session consists in many short trials (in this case 30 seconds) and is performed on consecutive days. Notwithstanding, if this design is adopted, it must be further examined if it allows the effects to persist. Furthermore, this study targeted healthy subjects and the conclusions may not apply in the case of clinical applications. It is considered that the present work introduces new findings in the field, as to the best of the authors knowledge there are no other studies in which two groups with different training intensity modalities were compared or intensity parameters were exhaustively studied. In the future, we encourage the use of carefully detailed reports of training parameters, as this is a crucial step towards thoroughly studying this topic.

Conflict of Interest Statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
Author Contributions

IE was responsible for designing and performing the experiment, analysing the results and writing the manuscript. WN guided the analysis of the results and participated in the writing of the manuscript. AC also guided the analysis of the results. CA guided the design of the protocol and the experimental acquisitions. FM contributed with expertise to solve technical issues throughout the course of the experiment. AR supervised the experiment and the writing of the manuscript, providing essential feedback. All authors contributed to the general discussion and review of the manuscript.

Funding

This work was supported by the Portuguese Science Foundation (FCT) through Project I&D LARSys UID/EEA/50009/2013, funded by national funds through FCT/MCTES (PID-DAC), and Grant BL32/2018, IST-ID.

Acknowledgments

Part of this content first appeared in a Master thesis by Esteves (2017), and is archived at Instituto Superior Técnico (University of Lisbon) website.

References

Angelakis, E., Stathopoulou, S., Frymiare, J. L., Green, D. L., Lubar, J. F., and Kounios, J. (2007). EEG Neurofeedback: A Brief Overview and an Example of Peak Alpha Frequency Training for Cognitive Enhancement in the Elderly. The Clinical Neuropsychologist, 21(1):110–129.

Arnold, L. E., Lofthouse, N., Hersch, S., Pan, X., Hurt, E., Bates, B., Kassouf, K., Moore, S., and Grantier, C. (2013). EEG Neurofeedback for ADHD: Double-Blind Sham-Controlled Randomized Pilot Feasibility Trial. Journal of Attention Disorders, 17(5):410–419.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society, 57(1):289–300.

Bolea, A. S. (2010). Neurofeedback Treatment of Chronic Inpatient Schizophrenia. Journal of Neurotherapy, 14(1):47–54.
Chernick, M. R. and Friis, R. H. (2003). *Introductory Biostatistics for the Health Sciences: Modern Applications Including Bootstrap*. Wiley Series in Probability and Statistics. Wiley-Interscience, Hoboken, N.J.

Conover, W. J. and Iman, R. L. (1981). Rank Transformations as a Bridge between Parametric and Nonparametric Statistics. *The American Statistician*, 35(3):124–129.

Corporation, R. H. (2017). 36-Item Short Form Survey (SF-36). https://www.rand.org/health/surveys_tools/mos/36-item-short-form.html.

Dekker, M. K., Sitskoorn, M. M., Denissen, A. J., and van Boxtel, G. J. (2014). The time-course of alpha neurofeedback training effects in healthy participants. *Biological Psychology*, 95:70–73.

Dias, A. M. and van Deusen, A. (2011). A New Neurofeedback Protocol for Depression. *The Spanish journal of psychology*, 14(01):374–384.

Escolano, C., Aguilar, M., and Minguez, J. (2011). EEG-based upper alpha neurofeedback training improves working memory performance. In *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE*, pages 2327–2330. IEEE.

Escolano, C., Navarro-Gil, M., Garcia-Campayo, J., Congedo, M., and Minguez, J. (2014). The Effects of Individual Upper Alpha Neurofeedback in ADHD: An Open-Label Pilot Study. *Applied Psychophysiology and Biofeedback*, 39(3-4):193–202.

Escolano, C., Olivan, B., Lopez-del-Hoyo, Y., Garcia-Campayo, J., and Minguez, J. (2012). Double-blind single-session neurofeedback training in upper-alpha for cognitive enhancement of healthy subjects. In *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE*, pages 4643–4647. IEEE.

Esteves, I. (2017). *Neurofeedback Training Intensity: Impact on EEG Self-Regulation*. Master Thesis, Instituto Superior Técnico, University of Lisbon, Lisbon.

Ferreira, P. L., Noronha Ferreira, L., and Nobre Pereira, L. (2012). Medidas sumário física e mental de estado de saúde para a população portuguesa. *Revista Portuguesa de Saúde Pública*, 30(2):163–171.

Gruzelier, J. H. (2014). EEG-neurofeedback for optimising performance. I: A review of cognitive and affective outcome in healthy participants. *Neuroscience & Biobehavioral Reviews*, 44:124–141.
Hsueh, J.-J., Chen, T.-S., Chen, J.-J., and Shaw, F.-Z. (2016). Neurofeedback training of EEG alpha rhythm enhances episodic and working memory: Alpha Neurofeedback on Memory. *Human Brain Mapping*, 37(7):2662–2675.

Kirchner, W. K. (1958). Age differences in short-term retention of rapidly changing information. *Journal of experimental psychology*, 55(4):352.

Klimesch, W. (1999). EEG alpha and theta oscillations reflect cognitive and memory performance: A review and analysis. *Brain research reviews*, 29(2):169–195.

Klimesch, W., Doppelmayr, M., Schwaiger, J., Auinger, P., and Winkler, T. (1999). Paradoxical alpha synchronization in a memory task. *Cognitive Brain Research*, 7(4):493–501.

Klimesch, W., Schack, B., and Sauseng, P. (2005). The Functional Significance of Theta and Upper Alpha Oscillations. *Experimental Psychology*, 52(2):99–108.

Kober, S. E., Schweiger, D., Reichert, J. L., Neuper, C., and Wood, G. (2017). Upper Alpha Based Neurofeedback Training in Chronic Stroke: Brain Plasticity Processes and Cognitive Effects. *Applied Psychophysiology and Biofeedback*, 42(1):69–83.

Lavy, Y., Dwolatzky, T., Kaplan, Z., Guez, J., and Todder, D. (2018). Neurofeedback Improves Memory and Peak Alpha Frequency in Individuals with Mild Cognitive Impairment. *Applied Psychophysiology and Biofeedback*.

Lecomte, G. and Juhel, J. (2011). The Effects of Neurofeedback Training on Memory Performance in Elderly Subjects. *Psychology*, 02(08):846–852.

Leins, U., Goth, G., Hinterberger, T., Klinger, C., Rumpf, N., and Strehl, U. (2007). Neurofeedback for Children with ADHD: A Comparison of SCP and Theta/Beta Protocols. *Applied Psychophysiology and Biofeedback*, 32(2):73–88.

Mayer, K., Blume, F., Wyckoff, S. N., Brokmeier, L. L., and Strehl, U. (2016). Neurofeedback of slow cortical potentials as a treatment for adults with Attention Deficit-/Hyperactivity Disorder. *Clinical Neurophysiology*, 127(2):1374–1386.

Meisel, V., Servera, M., Garcia-Banda, G., Cardo, E., and Moreno, I. (2013). Neurofeedback and standard pharmacological intervention in ADHD: A randomized controlled trial with six-month follow-up. *Biological Psychology*, 94(1):12–21.

Mirifar, A., Beckmann, J., and Ehrlenspiel, F. (2017). Neurofeedback as supplementary training for optimizing athletes’ performance: A systematic review with implications for future research. *Neuroscience & Biobehavioral Reviews*, 75:419–432.
MTA Cooperative Group (2004). National Institute of Mental Health Multimodal Treatment Study of ADHD Follow-up: Changes in Effectiveness and Growth After the End of Treatment. *PEDIATRICS*, 113(4):762–769.

Nan, W., Rodrigues, J. a. P., Ma, J., Qu, X., Wan, F., Mak, P.-I., Mak, P. U., Vai, M. I., and Rosa, A. (2012). Individual alpha neurofeedback training effect on short term memory. *International Journal of Psychophysiology*, 86(1):83–87.

Nan, W., Wan, F., Chang, L., Pun, S. H., Vai, M. I., and Rosa, A. (2017). An Exploratory Study of Intensive Neurofeedback Training for Schizophrenia. *Behavioural Neurology*, 2017:1–6.

National Collaborating Centre for Mental Health (London) (2009). *Attention Deficit Hyperactivity Disorder: Diagnosis and Management of ADHD in Children, Young People and Adults*. British Psychological Society, Leicester. OCLC: 731439170.

Navarro Gil, M., Escolano Marco, C., Montero-Marín, J., Minguez Zafra, J., Shonin, E., and García Campayo, J. (2018). Efficacy of Neurofeedback on the Increase of Mindfulness-Related Capacities in Healthy Individuals: A Controlled Trial. *Mindfulness*, 9(1):303–311.

Niv, S. (2013). Clinical efficacy and potential mechanisms of neurofeedback. *Personality and Individual Differences*, 54(6):676–686.

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. *Computational Intelligence and Neuroscience*, 2011:1–9.

Rockstroh, B., Elbert, T., Birbaumer, N., Wolf, P., Düchting-Röth, A., Reker, M., Daum, I., Lutzenberger, W., and Dichgans, J. (1993). Cortical self regulation in patients with epilepsy.pdf. *Epilepsy Research*, 1(14):63–72.

Rodrigues, J. a. P., Migotina, D. G., and da Rosa, A. C. (2010). EEG training platform: Improving Brain-Computer Interaction and cognitive skills. In *Human System Interactions (HSI), 2010 3rd Conference On*, pages 425–429. IEEE.

Rogala, J., Jurewicz, K., Paluch, K., Kublik, E., Cetnarski, R., and Wróbel, A. (2016). The Do’s and Don’ts of Neurofeedback Training: A Review of the Controlled Studies Using Healthy Adults. *Frontiers in Human Neuroscience*, 10.

Shapiro, S. S. and Wilk, M. B. (1965). An Analysis of Variance Test for Normality (Complete Samples). *Biometrika*, 52(3/4):591.
Siniatchkin, M., Hierundar, A., Kropp, P., Kuhnert, R., Gerber, W.-D., and Stephani, U. (2000). Self-regulation of slow cortical potentials in children with migraine: An exploratory study. *Applied psychophysiology and biofeedback*, 25(1):13–32.

Staufenbiel, S., Brouwer, A.-M., Keizer, A., and van Wouwe, N. (2014). Effect of beta and gamma neurofeedback on memory and intelligence in the elderly. *Biological Psychology*, 95:74–85.

Strehl, U., Aggensteiner, P., Wachtlin, D., Brandeis, D., Albrecht, B., Arana, M., Bach, C., Banaschewski, T., Bogen, T., Flaig-Röhr, A., Freitag, C. M., Fuchsenberger, Y., Gest, S., Gevensleben, H., Herde, L., Hohmann, S., Legenbauer, T., Marx, A.-M., Millenet, S., Pniewski, B., Rothenberger, A., Ruckes, C., Wörz, S., and Holtmann, M. (2017). Neurofeedback of Slow Cortical Potentials in Children with Attention-Deficit/Hyperactivity Disorder: A Multicenter Randomized Trial Controlling for Unspecific Effects. *Frontiers in Human Neuroscience*, 11.

Strehl, U., Leins, U., Goth, G., Klinger, C., Hinterberger, T., and Birbaumer, N. (2006a). Self-regulation of Slow Cortical Potentials: A New Treatment for Children With Attention-Deficit/Hyperactivity Disorder. *PEDIATRICS*, 118(5):e1530–e1540.

Strehl, U., Trevorrow, T., Veit, R., Hinterberger, T., Kotchoubey, B., Erb, M., and Birbaumer, N. (2006b). Deactivation of Brain Areas During Self-Regulation of Slow Cortical Potentials in Seizure Patients. *Applied Psychophysiology and Biofeedback*, 31(1):85–94.

To, E. Y., Abbott, K., Foster, D. S., and Helmer, D. (2016). Working Memory and Neurofeedback. *Applied Neuropsychology: Child*, 5(3):214–222.

Wan, F., Nan, W., Vai, M. I., and Rosa, A. (2014). Resting alpha activity predicts learning ability in alpha neurofeedback. *Frontiers in Human Neuroscience*, 8.

Wang, Y., Sokhadze, E. M., El-Baz, A. S., Li, X., Sears, L., Casanova, M. F., and Tasman, A. (2016). Relative Power of Specific EEG Bands and Their Ratios during Neurofeedback Training in Children with Autism Spectrum Disorder. *Frontiers in Human Neuroscience*, 9.

Welch, P. (1967). The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms. *IEEE Transactions on audio and electroacoustics*, 15(2):70–73.
Zoefel, B., Huster, R. J., and Herrmann, C. S. (2011). Neurofeedback training of the upper alpha frequency band in EEG improves cognitive performance. *NeuroImage, 54*(2):1427–1431.

Zola-Morgan, S. and Squire, L. R. (1993). Neuroanatomy of memory. *Annual review of neuroscience, 16*(1):547–563.

Table 1: Frequency ranges definitions

| Frequency Bands                  | Frequency Range          |
|---------------------------------|--------------------------|
| Individual Theta Band (ITB)     | 4 Hz to LTF              |
| Individual Lower Alpha Band 1 (LA1) | LTF to LTF + (IAF - LTF)/2 |
| Individual Lower Alpha Band 2 (LA2) | LTF + (IAF - LTF)/2 to IAF |
| Individual Upper Alpha Band (UA) | IAF to HTF               |
| Individual Alpha Band (IAB)     | LTF to HTF               |
| Sensorimotor Rhythm (SMR)       | 12 Hz to 15 Hz           |
| Beta Band (Beta)                | 16 Hz to 20 Hz           |
| Beta Band 2 (Beta2)             | 20 to 28 Hz              |
Table 2: Within group analysis across sessions: medians (Mdn), W statistic (W) and p-values (p) resulting from Wilcoxon signed rank test (right-tailed for UA band and two-tailed for the other frequency bands); medians with absolute value below 0.001 are shown as zero.

| Frequency bands | INTENSIVE | | SPARSE | | |
|-----------------|-----------|-----------------|-----------|-----------------|-----------|
|                 | A1        | A2              | A1        | A2              | |
|                 | Mdn       | W               | p         | Mdn             | W         | p         |
| ITB             | 0.028     | 29              | 0.496     | 0.028           | 31         | 0.359     |
| LA1             | -0.038    | 18              | 0.652     | -0.019          | 19         | 0.734     |
| LA2             | 0.034     | 25              | 0.820     | 0.014           | 25         | 0.820     |
| UA              | 0.011     | 24              | 0.455     | 0.003           | 27         | 0.326     |
| IAB             | 0.016     | 22              | 1.000     | 0.005           | 25         | 0.820     |
| SMR             | 0.009     | 34              | 0.203     | 0.010           | 36         | 0.129     |
| Beta            | -0.028    | 18              | 0.652     | -0.014          | 18         | 0.652     |
| Beta2           | -0.002    | 18              | 0.652     | -0.006          | 20         | 0.820     |

Table 3: Between groups analysis across sessions concerning the learning measures for different frequencies at the training location Fz: U statistic (U) and p-values (p) resulting from Wilcoxon rank sum Test.

| Frequency bands | A1 | A2 | | A1 | A2 | |
|-----------------|----|----|----|----|----|
|                 | U  | p  |    | U  | p  |    |
| ITB             | 36 | 0.730 | 42 | 0.931 | |
| LA1             | 21 | 0.094 | 27 | 0.258 | |
| LA2             | 29 | 0.340 | 45 | 0.730 | |
| UA              | 33 | 0.546 | 46 | 0.667 | |
| IAB             | 23 | 0.136 | 41 | 1.000 | |
| SMR             | 63 | 0.050 | 63 | 0.050 | |
| Beta            | 41 | 1.000 | 36 | 0.730 | |
| Beta2           | 46 | 0.667 | 42 | 0.931 | |
### Table 4: Within group analysis within session: median (Mdn), W statistic (W) and p-values (p) resulting from Wilcoxon signed rank test (right-tailed for UA band and two-tailed for the other frequency bands); medians with absolute value below 0.001 are shown as zero

| Frequency Bands | INTENSIVE | | | | SPARSE | | |
|-----------------|-----------|---|---|---|-----------|---|---|
|                 | W1 | W2 | | | W1 | W2 | |
|                 | W  | p  | | | Mdn | W  | p  | | | Mdn | W  | p  | |
| ITB             | 0.003 | 22 | 1.000 | | | 0.001 | 24 | 0.910 | | | -0.002 | 20 | 0.820 | |
| LA1             | 0.113 | 40 | 0.039 | | | 0.007 | 44 | 0.008 | | | -0.001 | 20 | 0.820 | |
| LA2             | 0.058 | 39 | 0.055 | | | 0.009 | 44 | 0.008 | | | -0.001 | 20 | 0.820 | |
| UA              | 0.036 | 40 | 0.020 | | | 0.006 | 43 | 0.006 | | | 0.003 | 17 | 0.752 | |
| IAB             | 0.067 | 45 | 0.004 | | | 0.008 | 45 | 0.004 | | | 0.002 | 21 | 0.910 | |
| SMR             | -0.005 | 21 | 0.910 | | | 0.001 | 20 | 0.820 | | | -0.001 | 22 | 1.000 | |
| Beta            | 0.018 | 25 | 0.820 | | | 0.009 | 36 | 0.129 | | | 0.003 | 38 | 0.074 | |
| Beta2           | -0.042 | 4 | 0.027 | | | -0.005 | 0 | 0.004 | | | -0.004 | 14 | 0.359 | |

### Table 5: Between groups analysis within session concerning the learning measures for different frequencies at the training location Fz: U statistic (U) and p-values (p) resulting from Wilcoxon rank sum test

| Frequency Bands | W1 | | | | W2 | | |
|-----------------| U  | p  | | | U  | p  | |
| ITB             | 37 | 0.796 | | | 38 | 0.863 | |
| LA1             | 65 | 0.032 | | | 45 | 0.730 | |
| LA2             | 63 | 0.050 | | | 55 | 0.222 | |
| UA              | 68 | 0.014 | | | 72 | 0.004 | |
| IAB             | 78 | <0.001 | | | 74 | 0.002 | |
| SMR             | 40 | 1.000 | | | 39 | 0.931 | |
| Beta            | 40 | 1.000 | | | 18 | 0.050 | |
| Beta2           | 17 | 0.040 | | | 24 | 0.162 | |
Figure 1: Top: Training diagram for the INTENSIVE group; Bottom: Training diagram for the SPARSE group

Figure 2: Top: Session diagram for the INTENSIVE group; Bottom: Session diagram for the SPARSE group
Figure 3: Left: Evolution across sessions for the INTENSIVE (blue) and SPARSE (red) groups during neurofeedback at Fz; median values of UA amplitude; error bars show median absolute deviation. Right: Boxplot with distribution of A1 and A2 measures for the INTENSIVE (blue) and SPARSE (red) groups; the + represents outliers.

Figure 4: Topographic distribution of median measures across sessions for the UA band.
Figure 5: Left: Evolution within session for the INTENSIVE (blue) and SPARSE (red) groups during neurofeedback at Fz; mean of UA amplitude in all corresponding blocks considering the median of the participants; errors bars show standard deviation. Right: Boxplot with distribution of W1 and W2 measures for the INTENSIVE (blue) and SPARSE (red) groups; the + represents outliers; the * represents p-values below 0.05 and the ** represents p-values below 0.01
Figure 6: Topographic distribution of median measures within session for the UA band
Figure 7: Comparison of relative amplitudes during pre-NF baseline, post-NF baseline, with EO and EC, and NF (using UA relative amplitudes that are not normalized with the baseline) for INTENSIVE (top) and SPARSE (bottom) groups; error bars show the median absolute deviation.

Figure 8: Boxplot with distribution of $t_{fu} - t_f$ UA amplitudes for the INTENSIVE (blue) and SPARSE (red) groups; the + represents outliers.