Identifying Indices of Learning for Alpha Neurofeedback Training

T. Dempster · D. Vernon

Abstract Neurofeedback has been around for decades and has applications for both clinical and healthy populations yet there is no standard method for measuring learning or a way of defining successful learning. Thus the aim of this study was to focus on alpha neurofeedback and examine changes in three different measures: amplitude, percent time, and integrated alpha, across four methods: within sessions, across sessions, within sessions compared to baseline, and across sessions compared to baseline. Participants completed 10 weekly sessions of eyes open alpha (8–12 Hz) neurofeedback training (NFT) at Pz. Whilst all three measures showed changes within sessions, the inclusion of baselines revealed that such changes represented a return to baseline levels rather than an increase in alpha. Changes across sessions were only evident in amplitude and inclusion of baseline showed that NFT did not elicit any changes beyond baseline levels. Given this a case is made for incorporating baseline measures when attempting to identify evidence of learning. It is also suggested that both amplitude and percent time measures are used independently rather than incorporate them into a more conservative and less sensitive integrated measure. Finally, focusing on within sessions changes may be a more useful approach in identifying changes resulting from NFT.

Keywords Alpha neurofeedback · Learning indices · Amplitude · Percent time · Baseline

T. Dempster (✉) · D. Vernon
Department of Applied Social Sciences (Psychology Department), Canterbury Christ Church University, North Holmes Road, Canterbury, Kent CT1 1QU, UK
e-mail: td31@canterbury.ac.uk

Introduction

Neurofeedback training (NFT) represents a sophisticated form of biofeedback based on specific aspects of cortical activity. The aim of such training is to teach the individual how to modify aspects of their own brain activity and in doing so potentially influence their behaviour. Such training typically involves recording an individual's electrical brain activity, as measured by the electroencephalograph (EEG), filtering this to isolate a target frequency component and then feeding specific information relating to this component back to the participant in the form of an audio and/or visual signal. For example, the goal of NFT may be to enhance the amplitude of a particular frequency component of the EEG, the level of which can be depicted using audio feedback, with higher amplitude related to higher volume or pitch of sound and lower amplitude to lower volume and/or pitch. The aim then is for the individual to recall and recreate states associated with periods when the amplitude/sound is higher, and in this way learn to associate specific changes in mental states with changes in brain activity.

NFT as a mechanism for altering brain activity and thereby influencing behaviour has been used with both clinical and healthy populations. For instance, it has gained widespread support as an intervention for those suffering from Attention Deficit Hyperactivity Disorder (ADHD) where it has been concluded that NFT represents a promising behavioural treatment (see Butnik 2005). In addition, it has been suggested to be of considerable clinical value for pathologies that demonstrate abnormal regulation of cortical arousal, such as epilepsy (see Sterman and Egner 2006). NFT has also been used to enhance a range of skills in healthy participants, including sporting performance, cognitive abilities, creativity and artistic performance (see

Vernon 2005; Vernon and Gruzelier 2008). The rationale for using NFT to enhance the performance of healthy participants is one based upon associations identified between cognitive processing and cortical activity (Vernon 2005). However, this idea has yet to be fully developed as research has shown that NFT can elicit changes beyond the target training frequency (see e.g., Egner et al. 2004).

Despite the fact that this technique has been available for a number of decades it is still unclear how to measure or identify evidence of learning when undergoing NFT. That is, what measures are used to identify whether an individual has been successful in altering the amplitude and/or frequency of a particular component of their EEG in the desired direction or not. Identifying an index, or indices, of learning incorporates two elements: how changes in the target frequency component of the EEG are measured and how successful learning is defined. Here we examine the effects of alpha NFT in an attempt to identify and clarify indices of learning. Though we focus on the alpha frequency component of the EEG the questions posed and the answers that emerge are relevant beyond the bounds of this frequency range and may apply to other components of the EEG.

A review of the literature reveals that the three most common measures used to identify changes in alpha activity are percent time (e.g., Angelakis et al. 2007; Cott et al. 1981a, 1981b; Yamaguchi 1980) amplitude (e.g., Cho et al. 2008; Putnam 2000) and integrated alpha (e.g., Knox 1982; Plotkin and Rice 1981; Tyson 1987). Percent time refers to the percentage of time participants spend above or below the target threshold when attempting to enhance and/or inhibit their alpha. Amplitude simply refers to changes in the mean level of amplitude during NFT, and integrated alpha refers to a measure that combines both the amount of time spent over/under a pre-set threshold and the amplitude (e.g., integrated alpha = percent time × amplitude/100).

To add to the confusion there are several possible methods of identifying possible changes in alpha using the above measures. These include, changes identified within the NFT session, changes across the NFT sessions, changes within sessions compared to a resting baseline and changes across sessions compared to a baseline. It should come as no surprise that given such a variety of measures and methods that some report changes in alpha as a function of NFT (e.g., Angelakis et al. 2007; Hanslmayr et al. 2005) whilst others do not (e.g., Gertz and Lavie 1983; Orenstein and McWilliams 1976) and that the differences in measures used has also led some to report changes within NFT sessions but not across sessions (DeGood and Valle 1978), whilst others find changes across sessions but not within (Cho et al. 2008). Furthermore, some find evidence of learning when incorporating a resting baseline (Hanslmayr et al. 2005) whilst others fail to include any baseline measures (e.g., Angelakis et al. 2007).

Thus, evidence of learning to alter alpha as a function of NFT seems to rely heavily on the type of measure used and the particular method employed. Unfortunately there is little consistency in the literature, which is problematic for a number of reasons. First, such a variety of measures and methods used makes it difficult to identify potentially useful trends in the data, which could hamper progress in the field. Second, such methodological variability makes comparisons across studies difficult, if not impossible. Whilst some have suggested that measures of learning should be standardised across studies (see e.g., Travis et al. 1975) there seems to be little evidence of this occurring. It is more often the case that researchers choose to focus on and report only one of the above measures without always making it clear why such a measure was selected.

Thus, the aim of this study was to examine alpha NFT using the three different measures across the four methods to see what, if any, differences emerge and whether this information can be used to identify an optimum measure of learning. In attempting to identify the most effective index/indices of learning the aim is to encourage other researchers within the field to adopt such a measure to encourage a more coherent and consistent approach to understanding the changes brought about by NFT.

Methods

Participants

Twenty-nine volunteers (6 male and 23 female, aged 18–44 years: $M = 21$) were screened and recruited from Canterbury Christ Church University. Of the 29 participants, 25 (5 male and 20 female) completed all 10 sessions of neurofeedback training, 2 (female) completed 7 sessions, 1 (female) completed 6, and 1 (male) completed 5. All participants had normal or corrected to normal vision and received a nominal payment of £40 for taking part (or proportionally less for the 4 who failed to complete all their sessions).

NFT

NFT was conducted over a period of 10 weeks, with each participant receiving one training session per-week. NFT was conducted using the ProComp Infiniti amplifier with Biograph 2.1 monitoring and feedback software (Thought Technology Ltd, Montreal, Quebec). Signal was acquired at 2,048 Hz, A/D converted and band filtered to extract the alpha (8–12 Hz) and EMG (40 Hz+) components. EEG was recorded from Pz with reference on the right and ground on the left earlobes respectively. Artefact rejection thresholds were set for each participant individually, suspending

feedback when eye-movements or muscle activity caused gross fluctuations in the EEG. Information relating to alpha activity was fed back to participants in either a visual, i.e. a moving bar on screen, and/or auditory format which depicted a changing tone. The greater the level, or amplitude, of alpha the higher the bar moved on the screen and the more intense the audio signal became. The aim was for participants to enhance their level of alpha beyond a pre-set threshold using the feedback to guide them. Thresholds were set individually based on the average amplitude recorded with eyes open at rest prior to each training session.

Procedure

Prior to taking part in the study the feedback loop was explained to each participant, which involved instructions to 'let the feedback guide them into maximising their alpha activity'. Each participant then completed ten sessions of NFT aimed at enhancing alpha (8–12 Hz) at Pz. Throughout each NFT session participants remained seated with their eyes open during each of the four stages. In stage one a resting baseline recording was taken while participants sat with their eyes open for 3 min and this was used to identify and set the threshold for that day's training. Stage two involved 7.5 min of feedback training, with the participant attempting to enhance the level of his/her alpha over and above the pre-set threshold. Stage three consisted of a 30 s blink break, to allow the participant to rest and re-focus. The final stage consisted of another 7.5 min of feedback training aimed at enhancing alpha.

Results

Indices of learning were assessed by examining changes in the three measures (amplitude, percent time, and integrated alpha) within the training sessions, within training sessions compared to baseline, across training sessions, and across sessions compared to baseline. For all analyses a Greenhouse–Geisser correction was used if Mauchley's test of sphericity was significant and all pairwise comparisons of main effects were conducted using a Bonferroni adjustment to control for Type I errors.

Within Sessions

For the within sessions analysis each training session contained two NFT segments of 7.5 min, divided by a 30 s blink break, although for the purposes of analysis each segment was divided into five 1.5 min periods of training in order to look for changes throughout the segments as well from one segment to another. Thus, changes within the NFT session were examined using a 2 (Segment: *Seg1* vs. *Seg2*) × 5

Table 1 Means and standard deviations (SD) for amplitude, percent time and integrated alpha (Iα) across the ten time periods within the NFT training sessions

Period	Mean amplitude	SD	Mean % time	SD	Mean Iα	SD
1	7.88	2.58	36.95	8.85	2.97	1.44
2	8.47	2.99	41.41	8.71	3.61	1.88
3	8.63	3.11	42.85	9.05	3.82	2.07
4	8.62	3.16	42.36	8.69	3.78	2.12
5	8.62	3.20	41.92	9.13	3.73	2.13
6	8.16	3.11	38.14	10.77	3.27	2.15
7	8.67	3.35	42.68	9.89	3.87	2.45
8	8.76	3.23	43.39	8.93	3.94	2.22
9	8.81	3.26	43.81	9.55	3.99	2.28
10	8.73	3.00	43.85	8.60	3.91	1.96

(Period: 1–5) repeated measures analysis of variance (ANOVA). The focus here was on changes in amplitude, percent time over threshold and integrated alpha scores across the ten training periods, see Table 1 for means and SD.

Amplitude

Analysis of amplitude showed a main effect of *Period* $F(2,63) = 24.75, p < 0.001, \eta^2 = 0.47$. However, there was no main effect of *Segment* $F(1,28) = 3.06$, $p = 0.09$, $\eta^2 = 0.10$ and no *Period* by *Segment* interaction $F(2,54) = 0.47$, $p = 0.62$, partial $\eta^2 = 0.02$. Pairwise comparisons of the main effect of *Period* revealed that amplitude was higher in *periods* 2 ($p < 0.001, d = 0.18$), 3 ($p < 0.001, d = 0.23$), 4 ($p < 0.001$, $d = 0.23$) and 5 ($p < 0.001$, $d = 0.22$), compared to *period* 1.

Percent Time

For percent time there was a main effect of *Period* $F(2,61) = 29.54$, $p < 0.001$, partial $\eta^2 = 0.51$, a marginal effect of *Segment* $F(1,28) = 3.87$, $p = 0.059$, partial $\eta^2 = 0.12$, but no *Period* by *Segment* interaction $F(3,75) = 0.70$, $p = 0.54$, partial $\eta^2 = 0.02$. Pairwise comparisons of the main effect of *Period* revealed that participants spent more time over their threshold in *periods* 2 ($p < 0.001, d = 0.49$), 3 ($p < 0.001, d = 0.61$), 4 ($p < 0.001, d = 0.61$) and 5 ($p < 0.001, d = 0.59$), compared to *period* 1. The marginal effect of *Segment* was due to participants spending more time over their threshold in *segment_2* compared to *segment_1* (42.3 and 41.1% respectively).

Integrated Alpha

Analysis of integrated alpha produced a main effect of *Period* $F(2,64) = 23.44$, $p < 0.001$, partial $\eta^2 = 0.46$.

However, there was no main effect of *Segment* $F(1,28) = 3.21$, $p = 0.08$, partial $\eta^2 = 0.10$ and no *Period* by *Segment* interaction $F(2,51) = 0.42$, $p = 0.64$, partial $\eta^2 = 0.02$. Pairwise comparisons of the main effect of *Period* showed that integrated alpha was higher during *periods* 2 ($p < 0.001$, $d = 0.31$), 3 ($p < 0.001$, $d = 0.38$), 4 ($p < 0.001$, $d = 0.38$) and 5 ($p < 0.001$, $d = 0.36$), compared to *period* 1.

Within Sessions Compared to Baseline

When examining the three measures within sessions compared to baseline it was not possible to add baseline as a factor as this was a single measure which remained constant across the session. Thus, for the within sessions analysis compared to baseline each measure was subtracted from its baseline to obtain a 'computed score' (e.g., mean amplitude during NFT − mean amplitude during resting baseline) and a 2 (Segment: *Segment_1* vs. *Segment_2*) × 5 (Period: 1–5) repeated measures ANOVA was conducted on this, see Table 2 for means and SD of the computed scores.

Amplitude

Analysis of computed scores for amplitude showed a main effect of *Period* $F(2,63) = 24.81$, $p < 0.001$, partial $\eta^2 = 0.47$, but no main effect of *Segment* $F(1,28) = 3.09$, $p = 0.09$, partial $\eta^2 = 0.10$ and no *Period* by *Segment* interaction $F(2,54) = 0.47$, $p = 0.62$, partial $\eta^2 = 0.02$. Pairwise comparisons of the main effect of *Period* revealed that the computed score was significantly lower at *period* 1 compared to *period* 2 ($p < 0.001$, $d = 0.38$), 3 ($p < 0.001$, $d = 0.47$), 4 ($p < 0.001$, $d = 0.47$) and 5 ($p < 0.001$, $d = 0.45$), see Fig. 1a. Additional one sample t tests comparing these computed scores to zero found no significant effects (all $p > 0.1$).
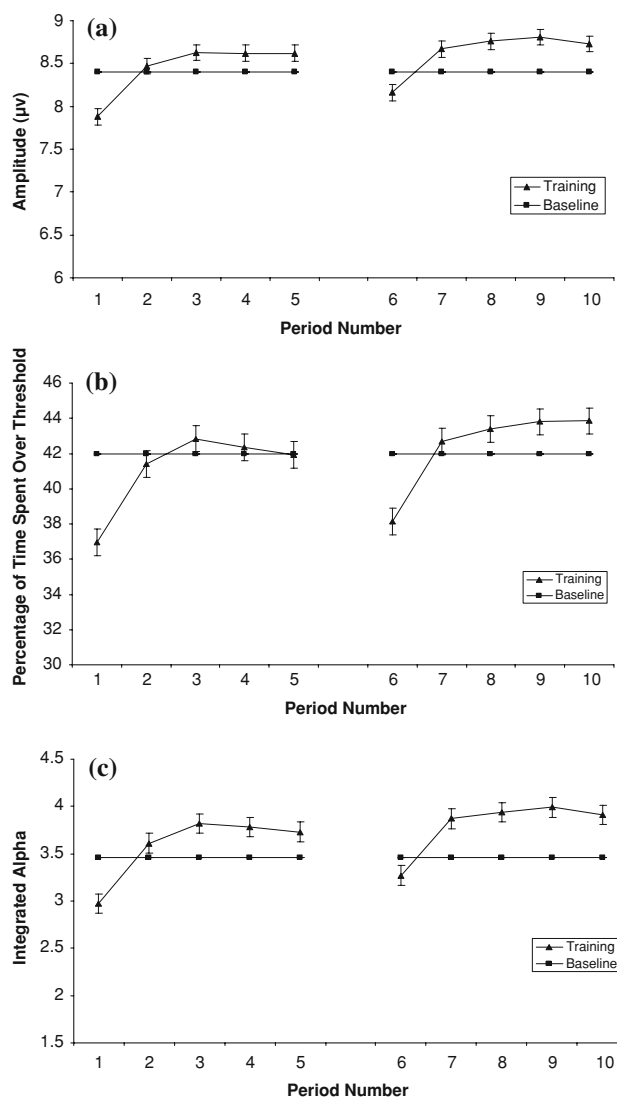

Fig. 1 Showing mean amplitude (**a**), percent time (**b**) and integrated alpha (**c**) (with standard *error bars*) during each period within sessions compared to baseline

Percent Time

For computed scores relating to percent time there was a main effect of *Period* $F(2,61) = 29.52$, $p < 0.001$, partial $\eta^2 = 0.51$, a marginal effect of *Segment* $F(1,28) = 3.88$, $p = 0.06$, partial $\eta^2 = 0.12$, but no *Period* by *Segment* interaction $F(3,75) = 0.71$, $p = 0.53$, partial $\eta^2 = 0.03$. Pairwise comparisons of the main effect of *Period* revealed that the computed score was significantly lower during *period* 1 compared to *periods* 2 ($p < 0.001$, $d = 0.45$), 3 ($p < 0.001$, $d = 0.56$), 4 ($p < 0.001$, $d = 0.55$) and 5 ($p < 0.001$, $d = 0.55$), see Fig. 1b. A one sample t test focusing on *period* 1 also showed this score to be significantly greater than zero (mean -4.46) $t(28) = -2.42$, $p < 0.05$, $d = 0.45$. The marginal effect of segment was

Table 2 Mean computed scores and standard deviation (SD) for amplitude (Amp_c), percent time (Time_c) and integrated alpha (Iα_c) across the ten time periods within the NFT training sessions

Period	Mean Amp_c	SD	Mean Time_c	SD	Mean Iα_c	SD
1	−0.52	1.20	−5.06	9.33	−0.49	1.28
2	0.07	1.39	−0.59	9.72	0.14	1.57
3	0.23	1.48	0.85	10.14	0.36	1.71
4	0.23	1.51	0.35	9.83	0.32	1.73
5	0.22	1.70	−0.08	9.76	0.27	1.81
6	−0.24	1.72	−3.87	11.40	−0.19	1.91
7	0.27	1.76	0.68	10.81	0.41	2.09
8	0.37	1.60	1.37	9.92	0.48	1.83
9	0.41	1.76	1.81	10.60	0.53	1.98
10	0.33	1.40	1.85	9.47	0.45	1.64

due to a lower overall score in *segment_1* compared to *segment_2* (−0.91 and 0.37 respectively).

Integrated Alpha

Analysis of the computed scores for integrated alpha produced a main effect of *Period* $F(2,64) = 23.55$, $p < 0.001$, partial $\eta^2 = 0.46$. There was no main effect of *Segment* $F(1,28) = 3.22$, $p = 0.08$, partial $\eta^2 = 0.10$, and no *Period* by *Segment* interaction $F(2,51) = 0.42$, $p = 0.64$, partial $\eta^2 = 0.02$. Pairwise comparisons of the main effect of *Period* showed the computed score was significantly lower at *period* 1 compared to *periods* 2 ($p < 0.001$, $d = 0.36$), 3 ($p < 0.001$, $d = 0.45$), 4 ($p < 0.001$, $d = 0.44$) and 5 ($p < 0.001$, $d = 0.42$), see Fig. 1c. Additional one sample t tests comparing these scores to zero found no significant effects (all $p > 0.1$).

Across Sessions

For the across session analysis each of the three measures was examined using a one-way ANOVA with 'Session' representing a single factor with ten levels, see Table 3 for means and SD.

Amplitude

A one way ANOVA revealed a main effect of *Session* $F(4,97) = 5.03$, $p < 0.01$, partial $\eta^2 = 0.19$. Pairwise comparisons showed that mean amplitude was higher at *session* 9 compared to *sessions* 1 ($p < 0.05$, $d = 0.46$) and 3 ($p < 0.05$, $d = 0.31$), there was also a marginal difference between *session* 10 and *session* 1, with amplitude at *session* 10 being marginally higher ($p = 0.07$, $d = 0.39$).

Percent Time

For percent time a one way ANOVA found no main effect of *Session* $F(5,113) = 0.75$, $p = 0.59$, partial $\eta^2 = 0.03$.

Integrated Alpha

There was no main effect of *Session* $F(4,91) = 2.13$, $p = 0.08$, partial $\eta^2 = 0.09$ for integrated alpha.

Across Sessions Compared to Baseline

Comparing the measures to baseline involved a 2 (Enhance: *Training* vs. *Baseline*) × 10 (Session: 1–10) repeated measures ANOVA. The means and standard deviations for the measures across the ten sessions are given in Table 3 whilst the baseline measures are given in Table 4.

Table 3 Means and standard deviations (SD) for amplitude, percent time and integrated alpha (Iα) across the ten NFT training sessions

Session	Mean amplitude	SD	Mean % time	SD	Mean Iα	SD
1	7.64	2.56	41.12	9.34	3.19	1.43
2	7.90	2.92	41.71	9.95	3.38	1.74
3	8.02	3.18	41.94	9.44	3.52	2.16
4	8.09	3.38	40.29	14.03	3.55	2.83
5	8.60	3.63	42.53	11.70	3.91	2.74
6	8.87	4.02	43.64	11.10	4.10	2.89
7	8.46	3.34	41.06	13.51	3.70	2.73
8	8.69	3.59	41.87	11.57	3.80	2.31
9	9.02	3.59	44.59	9.84	4.17	2.44
10	8.81	3.47	44.07	10.05	4.00	2.40

Amplitude

There was a significant main effect of *Session* $F(5,112) = 4.60$, $p < 0.01$, partial $\eta^2 = 0.17$, but no main effect of *Enhance* $F(1,22) = 0.85$, $p = 0.37$, partial $\eta^2 = 0.04$, and no *Session* by *Enhance* interaction $F(5,100) = 0.68$, $p = 0.63$, partial $\eta^2 = 0.03$. Pairwise comparisons of the main effect of *Session* showed that amplitude was higher at *session* 10 compared to *session* 1 ($p < 0.05$, $d = 0.33$), higher at *session* 9 compared to *session* 3 ($p < 0.05$, $d = 0.30$) and marginally higher at *session* 9 compared to *session* 1 ($p = 0.06$, $d = 0.38$), see Fig. 2a.

Percent Time

There was no main effect of *Session* $F(5,103) = 1.34$, $p = 0.25$, partial $\eta^2 = 0.06$, no main effect of *Enhance* $F(1,21) = 0.07$, $p = 0.94$, partial $\eta^2 = 0.00$, and no *Session* by *Enhance* interaction $F(5,108) = 0.71$, $p = 0.62$, partial $\eta^2 = 0.03$, see Fig. 2b.

Integrated Alpha

There was a significant main effect of *Session* $F(5,101) = 3.29$, $p < 0.01$, partial $\eta^2 = 0.14$, but no main effect of *Enhance* $F(1,21) = 0.21$, $p = 0.65$, partial $\eta^2 = 0.01$, and no *Session* by *Enhance* interaction $F(5,97) = 0.61$, $p = 0.68$, partial $\eta^2 = 0.03$, see Fig. 2c. Pairwise comparisons of the main effect of *Session* showed that integrated alpha was higher at *session* 9 compared to *session* 3 ($p < 0.05$, $d = 0.35$) and marginally higher at *session* 10 compared to *session* 1 ($p = 0.06$, $d = 0.27$).

Discussion

Examination of the three measures within the NFT sessions produced a consistent pattern with all showing an increase

Table 4 Means and standard deviations (SD) for baseline measures of amplitude, percent time and integrated alpha (Iα) across the ten NFT sessions

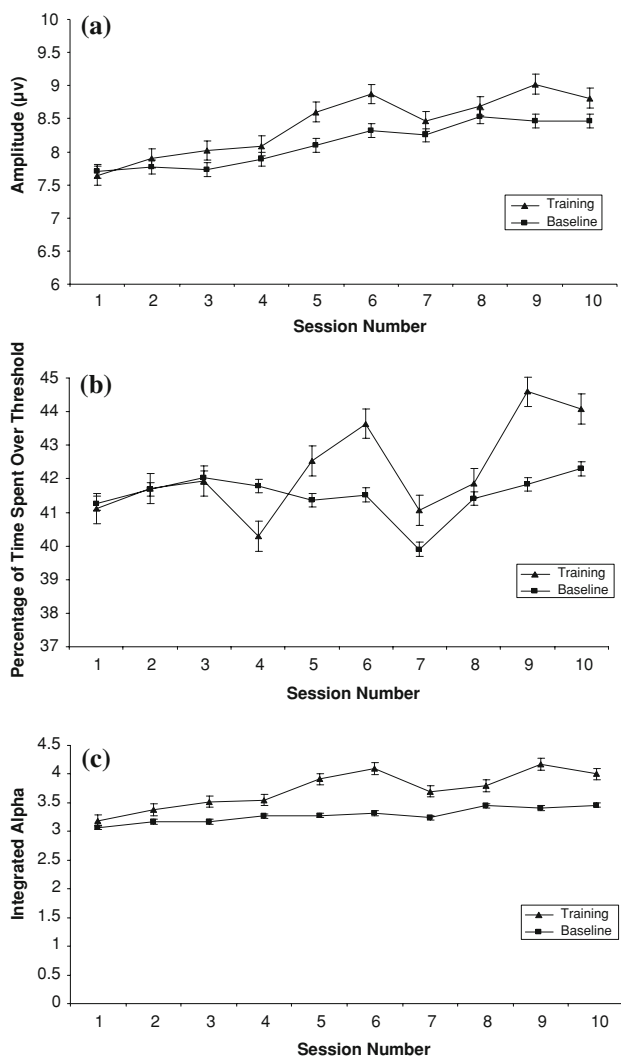| Session	Mean Amp_Baseline	SD	Mean %Time_Baseline	SD	Mean Iα_Baseline	SD
1	7.70	2.72	41.27	2.44	3.07	1.04
2	7.77	2.86	41.69	2.42	3.17	1.27
3	7.73	2.37	42.03	2.05	3.17	0.88
4	7.89	2.33	41.78	2.69	3.27	1.00
5	8.10	2.52	41.36	2.58	3.28	1.03
6	8.32	3.08	41.52	2.31	3.32	1.11
7	8.25	2.62	39.90	3.92	3.24	1.04
8	8.53	3.02	41.41	4.27	3.45	1.26
9	8.47	2.58	41.84	2.66	3.41	0.89
10	8.47	3.03	42.30	2.17	3.46	1.17

Fig. 2 Mean amplitude (**a**), percent time (**b**) and integrated alpha (**c**) (with standard *error bars*) during NFT compared to each session's corresponding baseline

from the beginning of the session to the end. However, when each of the three measures was compared to its relevant baseline it emerged that this pattern was due to scores

at the start of the training session being below the baseline and that over time these scores merely reverted to baseline levels.

For the across sessions data an increase was found for amplitude only, no change was evident in percentage time or integrated alpha. Furthermore, when the three measures were compared to baseline no differences emerged between measures obtained during NFT and those taken at rest, nevertheless, measures of both amplitude and integrated alpha increased over time when baseline alpha was incorporated into the analyses.

The initial analysis within sessions using the three different measures, if taken at face value, would suggest that participants were able to learn to alter their alpha activity, enhancing it in the desired direction from the start of the training session to the end. Such a finding is consistent with others who have reported changes within sessions without making any reference to baseline levels of activity (De-Good and Valle 1978; Fell et al. 2002; Putnam 2000; Tyson 1987). However, when training scores were subtracted from a relevant baseline measure it was possible to see that this difference was in fact due to a large decrease occurring in the level of alpha during period 1 which over time returned to baseline levels. Thus, instead of exhibiting evidence of learning to enhance their alpha activity using NFT it seems that undergoing such training led to an initial decrease in alpha which slowly, throughout the training session, returned to baseline levels. A similar pattern was reported by Plotkin (1978) when conducting NFT using an eyes-closed paradigm. There may be a number of reasons why alpha activity exhibited an initial decline. For instance, Fell et al. (2002) suggest that the newness of a situation and/or undertaking a new task can have a suppressing effect on alpha activity. A similar suggestion was also put forward by Paskewitz and Orne (1973). An alternative possibility is that initial exposure to the feedback stimuli may have a suppressing effect on alpha and that over time participants simply habituate to the stimuli and their alpha activity reverts back to baseline levels (see

Lynch et al. 1974; Plotkin 1978). Of course it may also be a combination of the two however it is not possible here to differentiate between these alternatives. Nevertheless, given that participants completed ten sessions of NFT these possibilities would predict that the suppression of alpha seen during the initial NFT period would lessen across the sessions as participants become more familiar with the situation and feedback stimulus. However, there was no evidence that the suppression effect seen during the first period ameliorated over time.

A plausible alternative as to why alpha activity was suppressed at the start of training relates to the potential focus of the participant's attention when trying to produce the feedback. For instance, participants may begin their training with the intention of trying to elicit a particular response from the computer, such as the audio feedback. This may lead them to initially focus their attention externally, on the feedback information. Such an external focus of attention could lead to a decrease in alpha activity, consistent with the findings from Aftanas and Golocheikine (2001). In contrast, as the session proceeds their focus of attention is likely to turn inwards, in an attempt to identify states and feelings associated with the feedback in order to identify possible strategies that can be adopted. An increasingly inward focus has been shown to elicit greater levels of alpha activity (Aftanas and Golocheikine 2001; Cooper et al. 2003). Thus, at the start of each training session participants may be more externally focused, waiting for the desired feedback to occur indicating they are doing the 'right' thing and/or correctly utilising a previously successful strategy, and that once the feedback begins it reassures them that they are getting the desired results so their sense of anticipation, waiting for that initial validating reinforcement, disappears and they become less externally focused on the feedback itself and more internally focused on whatever it is they are doing in order to continue eliciting feedback.

The patterns of data from the within sessions and the within sessions compared to baseline analyses are especially intriguing as the former could be used to indicate evidence of learning to enhance alpha within sessions whilst the latter clearly does not. However, it should be stressed that the aim here is not to critique the effectiveness of NFT per se, but simply to try and identify a useful index, or indices, of learning. Nevertheless, when attempting to identify evidence of learning to alter cortical activity via NFT it would seem essential to compare such activity during NFT to a resting baseline as this may throw a different light on the pattern of data and eliminate the possibility of confusing evidence of learning to enhance amplitude using NFT with an initial suppression effect resulting from the completion of an unfamiliar task, or as a result of a change in focus of attention from external to internal. Thus, we would echo the call of others who have

suggested that baselines need to be incorporated when attempting to identify evidence of learning via NFT in order to see if the differences that occur are in the predicted direction (Kondo et al. 1979; Plotkin 1978; Vernon 2005).

Analysis of the data within sessions, both without and with a comparison to baseline, produced a similar pattern from all three measures. Thus, it seems, there is little to discriminate between the various measures when looking for evidence of learning within the training sessions. However, when looking for evidence of learning across the NFT sessions without a comparison to baseline a different pattern emerges from between the three measures. Enhanced alpha activity is only evident when looking at changes in amplitude. There was no change in percent time or integrated alpha. To some extent this is consistent with others who have found changes in amplitude across sessions but not when measured using percent time (see e.g., Cho et al. 2008). Such a pattern may be due to certain differences between amplitude and percent time. For instance, Hardt and Kamiya (1976) point out that the use of percent time may be less sensitive to changes than amplitude. They suggest that participants who exhibit brief increases in the amplitude of their alpha rhythm, in line with the aim of the NFT, may be showing evidence of learning, but because such changes are brief the amount of time spent over the threshold may be insufficient for them to be classified as learners when looking only at percent time. Furthermore, given the dynamic nature of the brain one may expect to see brief fluctuations in amplitude as opposed to a simple increase which remains stable over time.

Nevertheless, it is important not to fall into the trap of thinking that percent time provides little or no useful information that cannot be obtained simply by looking at amplitude alone. For example, if a participant learns to keep alpha activity consistently above a pre-set threshold for the majority of the training session, as opposed to the brief but high fluctuations that would be picked up by measuring changes in amplitude, then arguably they have learned to enhance the amplitude of alpha in the desired direction. If this change represents a 90% increase in the amount of time spent over threshold but that increase is only 0.1 µV above threshold then it is likely that a significant change would be seen in the percent time measure but not necessarily in the amplitude measure. In this instance percent time would be more sensitive to such changes compared to amplitude and as such including details from both as two separate measures may provide a clearer picture of the changes elicited during training than amplitude alone.

Furthermore, whilst a change in the EEG may be evident when looking at 'either' amplitude or percent time, changes in behaviour may 'only' be evident with one. For instance, it could be the case that enhancing amplitude elicits a distinct

effect on behaviour and/or cognition compared to enhancing percent time. This then raises the question of which measure to use as an 'index of learning' for NFT. We are aware of no research directly addressing this issue and as such it remains the domain of future researchers. Nevertheless, it is worth noting that only by reporting *both* amplitude and percent time would it be possible to explore the possible differential link between changes in one aspect of the EEG and changes in behaviour. Thus, it would seem pertinent to include information about both measures in order to investigate whether or not training based on changes in amplitude elicits greater changes in behaviour compared to training based on percent time.

It could of course be argued that a measure which shows positive results when other measures do not may be doing so because it is more susceptible to Type 1 errors. Whilst this is not what we believe to be occurring here, reporting both percent time and amplitude as two separate measures would be more likely to detect such an anomaly. If one of the two measures consistently shows positive results when the other does not it would highlight the need for further exploration as to why such a pattern was occurring. By only reporting one of the measures or by integrating them together into one measure possible discrepancies between them are likely to be missed.

One further reason for the difference between amplitude and percent time may also be related to the type of feedback given. For example, here feedback was given in the form of a moving bar which changed colour when it breached a pre-set threshold. In addition, when the bar exceeded the threshold a tone also sounded and the frequency of this tone changed as the bar continued to move up or down above the threshold. In this instance the moving bar represents changes in amplitude, the changing colour of the bar represents time spent in alpha and the tone provides information on both amplitude and time. Notwithstanding the argument that the absence of feedback may itself be construed as feedback, it may be that participants received more information regarding changes in amplitude than they did concerning the amount of time spent above threshold. In particular, any changes occurring below threshold would only be shown as changes in amplitude. This may have made it easier for participants to alter the amplitude of their alpha activity without the need to increase the time spent in alpha. Such a suggestion is consistent with the notion that changes based on continuous feedback may provide more information (Hardt and Kamiya 1976). However, it should be noted that this represents a speculative possibility and we are unaware of any data directly addressing the question of whether feedback based on amplitude alone compared to an equal amount of feedback based on time spent above threshold would be more effective at inducing changes in alpha during NFT.

The fact that changes in amplitude across NFT sessions when not compared to baseline were not also seen in the integrated measure would suggest that integrated alpha represents a less sensitive and/or more conservative estimate of change. This is consistent with the view put forward by Travis et al. (1975) who noted that integrated alpha measures appear to be less sensitive to small variations. Use of a more conservative measure that fails to identify changes occurring that are clearly visible when another measure is used represents a potentially limiting factor. A measure that is too conservative may fail to identify useful patterns and trends in the data. In addition, integrated alpha may hide information because if it shows a change it fails to reveal whether that change is also evident in amplitude and/or percent time (as with the across sessions compared to baseline data). Likewise, the integrated measure could indicate no change but fail to clarify whether amplitude and percent time both agree or if, as was the case with the across sessions data, this was the case for only one measure. Furthermore, as mentioned above, no empirical case has yet been made as to which is a more valid measure, amplitude or percent time, therefore we believe that it is prudent to include information about both. Although integrated alpha can do this it does so in a more conservative manner and fails to provide information about the two measures independently. Thus, it would seem far more informative to discuss amplitude and percent time separately rather than combining them to form one integrated measure.

It is worth noting that in the across sessions comparison to baseline analyses, the inclusion of a baseline led to a significant main effect of session for the integrated alpha measure. However, this should not be taken to indicate that the integrated alpha measure increased in sensitivity, rather such an effect is the result of including the baseline measures which showed a small increment over time (see Table 4). Such increases are likely to be the result of changes in amplitude.

The change in amplitude across sessions was seen for both the NFT and the baselines. This would suggest that NFT can lead to increased alpha amplitude over time, although such increases do not exceed a comparative baseline. It is worth noting that participants were not monitored as to what they were doing during their resting baselines and it is plausible that whatever they did to pass the time, e.g. letting their mind wander, may have been more conducive to enhancing their alpha amplitude than what they were doing during the NFT itself. The issue of what constitutes an appropriate baseline is one that has received some discussion (see e.g., Plotkin 1976) but has yet to be resolved. Until the issue has been resolved it might be worth questioning participants to ascertain what, if anything, they are doing during the recording of their baseline to see if any patterns emerge.

With regards to the increase in baseline amplitude across sessions there may be several explanations. One is based upon Fell et al.'s (2002) suggestion, as mentioned above, that the novelty of the situation may have a suppressing effect on alpha. Therefore it would be expected that baselines taken in the initial sessions, when participants are unfamiliar with the experimental context, are suppressed in comparison to later sessions when participants have habituated to the situation. Alternatively, the increase in baseline amplitude across sessions could be the result of a change in focus from external to internal events, keeping in mind that baseline measures were taken with eyes open. Initially, when participants attend feedback sessions in the lab and are attached to sensors their focus may be externally orientated, looking around the room, taking in their surroundings etc. However, as their surroundings become more familiar with each session their attention during baselines recordings may move to become more internally focused, daydreaming, letting their mind wander etc., and this transition from external to internal sources to occupy their mind whilst they pass their time may be what results in the increase in alpha amplitude across the sessions (see e.g., Cooper et al. 2003).

Another possibility is that baseline amplitude may be increasing alongside the increase in amplitude seen during NFT because such training may leave a residual effect on participants' resting alpha. In other words, consciously attempting to enhance alpha amplitude during NFT may result in greater levels of alpha amplitude beyond the training session itself. For example, Cho et al. (2008) found that participants' alpha amplitude at the end of each weekly training session positively correlated with the level of alpha amplitude seen in the next session's baselines. That is, the higher the amplitude of their participants' alpha at the end of each session the higher the amplitude of their alpha during the eyes open baseline in the next session. The long-term effects of alpha NFT on the brain outside of the training sessions themselves are not yet known (Vernon 2008) but it could be that learning to consciously increase a particular component of cortical activity produces changes in that component which outlast the session itself. If this is the case then it means that trying to identify an index of learning by focusing on across sessions changes may not be the most effective approach. Possible changes across sessions due to NFT may be confounded by concurrent changes in baselines. Thus, within sessions compared to baseline may represent a more effective method to use when looking for the evidence of learning to alter alpha amplitude via NFT because this provides a picture of the changes seen during the training session rather than the difference from one session to the next.

On a related point, it is of interest to note that previous research has indicated that NFT to enhance the amplitude of low beta (11.7–14.6 Hz) also produced changes in alpha

(7.8–11.7 Hz) (Egner et al. 2004). Whilst the focus of this article has been on enhancing the amplitude of alpha, such training can be conducted by focusing on the alpha frequency component in isolation from other components of the EEG (e.g., Cho et al. 2008), or by incorporating the surrounding frequencies of theta and beta into the NFT protocol (e.g. Dempster and Vernon 2008). It would be interesting to see if the indices of learning proposed here, amplitude and percent time, produce differential patterns and training effects when alpha amplitude is trained in isolation as compared to when it is trained alongside other frequencies.

Finally, as mentioned in the introduction, although we focus here on the alpha rhythm our findings and recommendations are also relevant to the other frequencies of the EEG. For example, Hoedlmoser et al. (2008) recently reported that ten sessions of NFT were sufficient to increase relative SMR (12–15 Hz) amplitude from 0.98 to 1.06 µV across sessions. However, a look at the baseline measures shows that these remained stable from the beginning of the training (1.09 µV) to the end (1.07 µV). As such, the reported increase in SMR over time is in fact due to a decrease in relative SMR during the early NFT sessions and that over time the amplitude merely reverts back to around baseline levels. A finding that mirrors very closely what we found when comparing within sessions learning to baseline.

In conclusion, we would argue that it is essential that baseline measures are included when attempting to identify evidence of learning via NFT. Incorporating a baseline measure not only controls for natural fluctuations but also enables the researcher to see whether any changes seen during NFT exceed the amount of alpha participants naturally produce, or if it merely reflects a return to participants' natural levels after an initially suppressing effect. Furthermore, given that both amplitude and percent time measure different aspects of the EEG it would seem prudent to include both measures but to look at them individually rather than combining them into a less sensitive measure. Furthermore, a combined measure also fails to provide a clear picture of where any changes occur, in amplitude, percent time, or both. Finally, focusing on changes within NFT sessions may be more fruitful as changes possible across sessions may be confounded by shifting baselines.

References

Aftanas, L. I., & Golocheikine, S. A. (2001). Human anterior and frontal midline theta and lower alpha reflect emotionally positive state and internalized attention: High resolution EEG investigation of meditation. *Neuroscience Letters, 310*, 57–60.

Angelakis, E., Stathopoulou, S., Frymiare, J. L., Green, D. L., Lubar, J. F., & 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*, 110–129.

Butnik, S. M. (2005). Neurofeedback in adolescents and adults with attention deficit hyperactivity disorder. *Journal of Clinical Psychology, 61*, 621–625.

Cho, M. K., Jang, H. S., Jeong, S.-H., Jang, I.-S., Choi, B.-J., & Lee, M.-G. T. (2008). Alpha neurofeedback improves the maintaining ability of alpha activity. *Neuroreport, 19*(3), 315–317.

Cooper, N. R., Croft, R. J., Dominey, S. J. J., Burgess, A. P., & Gruzelier, J. H. (2003). Paradox lost? Exploring the role of alpha oscillations during externally vs. internally directed attention and the implications for idling and inhibition hypotheses. *International Journal of Psychophysiology, 47*, 65–74.

Cott, A., Pavloski, R. P., & Black, A. H. (1981a). Operant conditioning and discrimination: Some methodological limitations inherent in response discrimination experiments. *Journal of Experimental Psychology: General, 110*(3), 398–414.

Cott, A., Pavloski, R. P., & Goldman, J. A. (1981b). Cortical alpha rhythm, biofeedback, and the determinants of subjective state. *Journal of Experimental Psychology: General, 110*(3), 381–397.

DeGood, D. E., & Valle, R. S. (1978). Self-reported alcohol and nicotine use and the ability to control occipital EEG in a biofeedback situation. *Addictive Behaviors, 3*, 13–18.

Dempster, T., & Vernon, D. (2008). The effect of distinct training schedules on participants' ability to alter their alpha activity via neurofeedback: A preliminary study. *Revista Espanola De Neuropsicologia, 10*(1), 176.

Egner, T., Zech, T. F., & Gruzelier, J. H. (2004). The effects of neurofeedback training on the spectral topography of the electroencephalogram. *Clinical Neurophysiology, 115*, 2452–2460.

Fell, J., Elfadil, H., Klaver, P., Röschke, J., Elger, C. E., & Fernández, G. (2002). Covariation of spectral and nonlinear EEG measures with alpha biofeedback. *International Journal of Neuroscience, 112*, 1047–1057.

Gertz, J., & Lavie, P. (1983). Biological rhythms in arousal indices: A potential confounding effect in EEG biofeedback. *Psychophysiology, 20*(6), 690–695.

Hanslmayr, S., Sauseng, P., Doppelmayr, M., Schabus, M., & Klimesch, W. (2005). Increasing upper alpha power by neurofeedback improves cognitive performance in human subjects. *Applied Psychophysiology and Biofeedback, 30*(1), 1–10.

Hardt, J. V., & Kamiya, J. (1976). Conflicting results in EEG alpha feedback studies: Why amplitude integration should replace percent time. *Biofeedback and Self-Regulation, 1*(1), 63–75.

Hoedlmoser, K., Pecherstorfer, T., Gruber, G., Anderer, P., Doppelmayr, M., Klimesch, W., et al. (2008). Instrumental conditioning of human sensorimotor rhythm (12–15 Hz) and its impact on sleep as well as declarative learning. *Sleep, 31*(10), 1401–1408.

Knox, S. S. (1982). Alpha enhancement, autonomic activation, and extraversion. *Biofeedback and Self-Regulation, 7*(4), 421–433.

Kondo, C. Y., Travis, T. A., Knott, J. R., & Bean, J. A. (1979). Effects of true and inverted feedback on integrated occipital alpha. *The Journal of Psychology, 102*, 101–106.

Lynch, J. J., Paskewitz, D. A., & Orne, M. T. (1974). Some factors in the feedback control of human alpha rhythm. *Psychosomatic Medicine, 36*(5), 399–410.

Orenstein, H. B., & McWilliams, B. (1976). Variations in electroencephalographic alpha activity under conditions of differential lighting and auditory feedback. *Biofeedback and Self-Regulation, 1*(4), 423–432.

Paskewitz, D. A., & Orne, M. T. (1973). Visual effects on alpha feedback training. *Science, 181*, 360–363.

Plotkin, W. B. (1976). On the self-regulation of the occipital alpha rhythm: Control strategies, states of consciousness, and the role of physiological feedback. *Journal of Experimental Psychology: General, 105*(1), 66–99.

Plotkin, W. B. (1978). Long-term eyes-closed alpha-enhancement training: Effects on alpha amplitudes and on experiential state. *Psychophysiology, 15*(1), 40–52.

Plotkin, W. B., & Rice, K. M. (1981). Biofeedback as a placebo: Anxiety reduction facilitated by training in either suppression or enhancement of alpha brainwaves. *Journal of Consulting and Clinical Psychology, 49*(4), 590–596.

Putnam, J. (2000). The effects of brief, eyes-open alpha brain wave training with audio and video relaxation induction of 77 army reservists. *Journal of Neurotherapy, 4*(1), 17–28.

Sterman, M. B., & Egner, T. (2006). Foundation and practice of neurofeedback for the treatment of epilepsy. *Applied Psychophysiology and Biofeedback, 31*(1), 21–35.

Travis, T. A., Kondo, C. Y., & Knott, J. R. (1975). Alpha enhancement research: A review. *Biological Psychiatry, 10*(1), 69–89.

Tyson, P. D. (1987). Task-related stress and EEG alpha biofeedback. *Biofeedback and Self-Regulation, 12*(2), 105–119.

Vernon, D. J. (2005). Can neurofeedback training enhance performance? An evaluation of the evidence with implications for future research. *Applied Psychophysiology and Biofeedback, 30*(4), 347–364.

Vernon, D. (2008). Neurofeedback: Using computer technology to alter brain functioning. In F. Orsucci & N. Sala (Eds.), *Reflexing interfaces: The complex coevolution of information technology ecosystems* (pp. 94–108). New York: IGI.

Vernon, D., & Gruzelier, J. (2008). Electroencephalographic biofeedback as a mechanism to alter mood, creativity and artistic performance. In B. N. DeLuca (Ed.), *Mind–body and relaxation research focus* (pp. 149–164). Nova Science.

Yamaguchi, H. (1980). Characteristics of alpha-enhancement biofeedback training with eyes closed. *Tohoku Psychologica Folia, 39*(1–4), 40–50.