**Methods**

*EEG frequency bands spectral features estimation*

The frequency bands analysis was carried out taking single days into account (i.e. daily wise), thus averaging sessions’ spectral features of each day and separately for each patient. The number of days for each patient was: F - 14; G - 17; B - 12; W - 6. The following frequency bands were taken into account for electroencephalographic (EEG) spectral features estimation: delta 0.25-3.5Hz, low-theta 3.5-5Hz, high-theta 5-8Hz, low-alpha 8-10Hz and high-alpha 10-13Hz. The first spectral feature was the mean power of each frequency band. Then, since no clear frequency peak changes were visible in the spectrums (in all the channels), the oscillatory activities were further investigated to ascertain the differences, if any, in variability of the frequency bands mean power. This required the computation of the second spectral feature, i.e. the standard deviation of the oscillatory activity, here called ‘frequency band power variability’, calculated day wise. Frequency bands mean power and its ‘variability’ were estimated using Welch’s method separately for: (i) the resting interval (i.e. 5s before the presentation of each true/false sentence), (ii) the sentence presentation (SP) interval and (iii) the inter-stimulus interval (ISI) corresponding to true and false sentences.

Three slightly different EEG pre-processing pipelines were used to compute spectral features [1,3-4,7]. They included: (i) no spatial filtering (mono-polar derivations), (ii) common-average-reference and (iii) bi-polar derivations. Because the three techniques yielded very similar results, only the results obtained applying the mono-polar derivations pipeline are presented herein for simplicity.

Further steps toward analysis of neuro-electric changes could be obtained computing temporal EEG dynamics by means of non-linear and time-frequency techniques. A preliminary step in this direction was carried out as described in the following paragraphs (i.e. *EEG temporal dynamics analysis*, *Session wise time-frequency data classification* and *EEG time-frequency and fNIRS classification comparison*).

*EEG middle frequency bands comparison*

Two separate hypotheses were tested: a) whether power spectral features of middle frequency bands (high-theta, low-alpha and high-alpha) differentiate "true/yes" and "false/no" sentences’ ISI and b) whether they differentiate the sentence presentation interval from ISI (regardless of "true/yes" and "false/no" conditions). For each patient, 2-way ANOVA was used to compare middle frequency bands features of "true/yes" and "false/no" sentences’ ISI (2 conditions and 6 channels) and 2-wayANOVA was used to compare middle frequency bands features of sentence presentation and inter-stimulus intervals (2 intervals and 6 channels).

*EEG low frequency bands correlation with fNIRS classification accuracy*

The hypothesis of whether low frequency bands were related to *f*NIRS classification accuracy was tested to find relevant relationships between the BCI experimental procedure outcome and the vigilance state of the patients. Initially Spearman’s correlation was used to evaluate the relationships between low frequency bands (delta, low-theta and high-theta) features and *f*NIRS classification accuracy. Then, averaged correlation between each frequency band mean power and *f*NIRS classification accuracy was computed across intervals (resting, sentence presentation and sentence’s ISI) and across all electrodes to test the hypothesis of whether the median of averaged correlation was zero (Wilcoxon signed rank test for zero median was used). Thus, for each patient the number of correlation coefficients averaged together was *Ix2x6*, with *I* number of intervals (from 1 to 3: resting, sentence presentation and sentence’s ISI), two conditions (true and false) and six channels (FC5, FC1, FC6, CP5, CP1 and CP6).

Successively, the number of days of each patient was split in successful and unsuccessful days according with the chance level threshold (i.e. days with classification accuracy above chance level threshold were considered successful). Then, if the hypothesis of the averaged correlation was rejected, the hypothesis of whether the particular band mean power distribution medians of successful and unsuccessful days differed was tested (Wilcoxon rank sum test for equal medians was used).

*EEG temporal dynamics analysis*

Continuous Wavelet Transform (CWT) [6] and Short-Time Fourier Transform (STFT) [5] features were computed to investigate EEG temporal dynamics of "true/yes" and "false/no" brain responses. Time-frequency features were extracted from averaged true/false sentences’ ISI of each session for each patient. EEG data were previously processed using the mono-polar derivations pre-processing pipeline. Due to recording issues and differences between the ISI length across patients, the length of averaged true/false sentences’ ISI of each session was fixed to 13s, 1s before ISI onset (i.e. 1s before the end of the true/false sentence) and 12s after ISI onset (i.e. 12s after the end of the true/false sentence). CWT setting properties were: Morlet wavelet, 24 scales (frequency range 1.7 - 40.6Hz), time resolution of 2ms. STFT setting properties were: frequency resolution of 1.95Hz (range 0 - 250Hz), time resolution of 64ms. For each patient, channel and session all time-frequency data of averaged true and false sentences’ ISI were used as input of an SVM-based classification procedure, separately for CWT and STFT features (see paragraph below). The purpose was the estimation of the true vs. false neuro-electric responses recognition accuracy. In parallel, the procedure served to localize relevant frequencies changes over the EEG time course of ISI between true and false conditions.

*Session wise time-frequency data classification*

The SVM-based classification procedure, applied separately for CWT and STFT features, comprised the following steps: (i) automatic selection of relevant true/false time-frequency features, across sessions and channels, by means of stepwise multi-linear regression [2]; (ii) randomized partition of the selected time-frequency features subset in two halves, one to train and the other one to test the SVM model. A 20-fold cross-validation procedure was applied to estimate the recognition accuracy of averaged true vs. false sentences’ ISI. Because a direct paired test between EEG time-frequency and *f*NIRS classification accuracy was not applicable (the former was performed at session level while the latter at single-trial level), the 20 time-frequency recognition accuracy estimates where compared to the mean value of *f*NIRS classification accuracy (using the t-test). In addition, the distribution of the automatically selected relevant time-frequency features was computed, again, separately for CWT and STFT.

**Results**

*EEG middle frequency bands analysis for "true/yes" and "false/no" sentences’ ISI*

Inspecting power spectrum density plots of each patient the dominant frequencies were identified. The patients showed the following stable dominant frequencies: F - 6.75 Hz, G - 6.25 Hz; B - 7 Hz; W - 8 Hz). The 2-way ANOVA used to compare the middle frequency bands (high-theta, low-alpha and high-alpha) mean power of "true/yes" and "false/no" sentences’ ISI (2 conditions and 6 channels) revealed no main effects of the conditions factor and the channels factor in any patient (all *p > 0.05*).

*EEG middle frequency bands analysis for sentence presentation and sentence’s ISI*

The 2-way ANOVA used to compare the middle frequency bands (high-theta, low-alpha and high-alpha) spectral features of sentence presentation and sentence’s ISI (2 intervals and 6 channels) revealed some main effects of the intervals factor (sentence presentation *vs.* sentence’s ISI) as reported in *S1 Table*, *Section A*. In two (G and B) out of four patients, the ANOVA of low-alpha band ‘power variability’ revealed a main effect of the intervals factor only (*p < 0.05*), meaning less low-alpha band ‘power variability’ in the sentence’s ISI compared to the sentence presentation interval (no main effect of channels factor was found). In patient W, the ANOVAs of high-theta, low-alpha and high-alpha bands mean power revealed a main effect of the intervals factor only (*p < 0.05*), meaning a higher high-theta, low-alpha and high-alpha bands mean power in the sentence’s ISI compared to the sentence presentation interval (sentence’s ISI was synchronized compared to sentence presentation interval). Again in this case, no main effect of channels factor was found. In patient F, no main effects of intervals and channels were found for each middle frequency band (all *p > 0.05*).

*Low EEG rhythms correlation with fNIRS classification accuracy results*

The correlation between *f*NIRS classification accuracy and low frequency bands (those more related to the vigilance) mean power revealed some interesting results (see *S1 Table*, *Section B*). In three (G, B and W) out of four patients the median of the negative averaged correlation between low-theta band mean power and *f*NIRS classification accuracy across intervals (resting, sentence presentation and sentence’s ISI) and across all electrodes was significantly different from zero (patient G: *r=-0.365*; patient B: *r=-0.264*; patient W: *r=-0.386*; all *p < 0.05*). In patient G and B this holds true even in delta band (patient G: *r=-0.198*; patient B: *r=-0.238*; all *p < 0.05*), while a significant positive averaged correlation in high-theta band for patient W was found (*r=0.313*; *p < 0.05*). While, in patient F the median of the positive averaged correlation between delta band mean power and *f*NIRS classification accuracy across intervals (resting, sentence presentation and sentence’s ISI) and across all electrodes was significantly different from zero (*r=0.233*; *p < 0.05*); in the same patient this holds true even in high-theta band (*r=0.213*; *p < 0.05*), while in low-theta band the averaged correlation was not significantly different from zero (*r=-0.017*; *p > 0.05*).

To corroborate these results, the low frequency bands mean power distribution medians of successful and unsuccessful days was further investigated for each patient to ascertain the difference, if any. In two out of four patients (G and B) the null hypothesis of equal median of successful and unsuccessful day was rejected (*p < 0.05*). In patient G, the low-theta band mean power distribution medians of successful and unsuccessful days differed significantly (low-theta band mean power of successful days was smaller than that of unsuccessful days). In patient B, the high-theta band mean power distribution medians of successful and unsuccessful days differed significantly (high-theta band mean power of successful days was smaller than that of unsuccessful days).

*EEG and EOG power spectrum density (PSD)*

*EEG time-frequency and fNIRS classification comparison*

The SVM-based classification procedure applied to the EEG time-frequency data of each patient yielded the results described in S2 Table.

The above results may only partially confirm the hypothesis of superiority of *f*NIRS over EEG in the detection of true and false neurophysiological responses to yes/no sentences: in two patients (i.e. F and W) this holds true; in patient B, CWT and STFT methods yielded different results (CWT comparable with *f*NIRS, STFT lower than *f*NIRS); in patient G, both CWT and STFT techniques proved a superior estimation than *f*NIRS. However, because CWT and STFT time-frequency analyses were carried out at session level (i.e. on averaged data, not on single-trial data) caution is necessary in the interpretation of the results. Therefore, a more complex and time-consuming analysis at single-trial level is necessary to confirm or disprove time-frequency observations on averaged data. Any further explanation of the here reported time-frequency results would have a speculative meaning.

*Additional references*

1. Blankertz B, Tomioka R, Lemm S, Kawanabe M, Müller KR. Optimizing Spatial filters for Robust EEG Single-Trial Analysis. IEEE Signal Processing Magazine. 2008; 25(1): 41-56.

2. Draper NR and Smith H. Applied Regression Analysis. John Wiley and Sons, 2nd Ed., New York, 1981.

3. Farina D, Jensen W, Akay M. Introduction to Neural Engineering for Motor Rehabilitation. Wiley-IEEE Press, 2013, 1st ed.

4. Farquhar J, Hill NJ. Interactions between pre-processing and classification methods for event-related-potential classification: best-practice guidelines for brain-computer interfacing. Neuroinformatics. 2013; 11(2): 175-92.

5. Oppenheim AV and Schafer RW. Discrete-Time Signal Processing. Prentice-Hall, Englewood Cliffs, NJ, 1989.

6. Mallat S. A Wavelet Tour of Signal Processing. Academic Press, 3rd Ed., Boston, 2009.

7. Wu W, Chen Z, Gao X, Li Y, Brown EN, Gao S. Probabilistic Common Spatial Patterns for Multichannel EEG Analysis. IEEE Trans Pattern Anal Mach Intell. 2015; 37(3): 639-53.