Decoding affective states across databases using functional near-infrared spectroscopy

Lucas R. Trambaiolli*[1], Juliana Tossato[1], André M. Cravo[1], Claudinei E. Biazoli Jr[1], João R. Sato[1]

[1]Mathematics, Computation and Cognition Center, Universidade Federal do ABC - Rua Santa Adélia, 166 - Santo André - SP, 09210-170, Brazil.

*Corresponding author: Lucas R. Trambaiolli
e-mail: lucasrtb@gmail.com
Work Phone: +55-11-96925-8585

Abstract: Multivariate brain decoding (MBD) can be applied to estimate mental states using brain signal measurements. In the best scenario, a MBD model should be trained in a first set of volunteers and then validated in a new and independent dataset. Here, we aimed to evaluate whether functional near-infrared spectroscopy (fNIRS) signals from frontal and occipital areas provide enough information to discriminate affective states. For this purpose, a linear discriminant analysis classifier was trained in a first database (49 participants, 24.65±3.23 years) and tested in an independent database (20 participants, 24.00±3.92 years). Significant accuracies were found for positive vs. negative (64.50±12.44%, $p<0.01$) and negative vs. neutral (67.75±14.45%, $p<0.01$) affect during a passive elicitation condition, consisting in viewing pre-validated images with emotional content. For the active elicitation condition, in which volunteers were instructed to recollect personal affective experiences, significant accuracy was found for positive vs. neutral affect (71.00±17.93%, $p<0.01$). In this last case, only three fNIRS channels were sufficient to discriminate between those affective states: two positioned over the left ventrolateral prefrontal area and one over the right lateral orbitofrontal cortex. In conclusion, our results show that fNIRS is a feasible technique for inter-subject affective decoding, reaching significant classification accuracies using a few and biologically consistent features.

Keywords: functional near infrared spectroscopy, affective decoding, multivariate analysis, affective states

1. INTRODUCTION

Multivariate brain decoding (MBD) allows the estimation of mental states based only on measurements of brain signals (Haynes and Rees 2006). In comparison to traditional analytical methods, this approach has the advantage of considering the various brain regions simultaneously, thus providing information of the neural networks of interest as whole (Kubilius et al. 2015). MBD applications vary from simple to quite complex processes, including the prediction of new visual scenes, auditory and visual imagery, semantic differentiation, affective processing and decision-making (for a review see Tong and Pratte 2012). Particularly concerning affective decoding, fMRI studies reached up to 92% of MBD accuracy in a single participant classification of

positive versus negative valence (Baucom et al. 2012). Considering these results, real-world applications of MBD become attractive. For instance, Lawrence et al. (2014) developed a neurofeedback system for self-regulation of the right anterior insula in a affective paradigm. After four training runs, volunteers presented a linear increase in activation of this area (Lawrence et al. 2014).

Ideally, a real-world MDB system for therapeutic application should be trained with a first set of volunteers and applied with little or no calibration to new and independent individuals. However, the inter-subject classification of mental states remains an open technical challenge in the MBD field (Haynes and Rees 2006). For affective states, inter-subject classifications have been previously reported on facial expressions and body language (Mariooryad and Busso 2013; Tan et al. 2016), voice tone (Mariooryad and Busso 2013). A critical limitation of those studies is that volunteers can easily handle the records used in order to achieve the desired result, even without changing the mental affective state. Moreover, such measures do not provide crucial neural information to the MBD system. Thus, investigations with neurophysiological records such as electrophysiology (EEG) or neuroimaging (e.g. fMRI and fNIRS) deserve further attention. Using EEG data, for example, some studies combined relevant attributes selection with clusters of classifiers to vote about the level of the valence of an individual, achieving accuracies between 70 and 95% (Bozhkov and Petia 2014, Georgieva et al. 2015, Ramaraju, Izzidien and Roula 2015). Using fMRI data, accuracy varied from 60% to 80% depending on the number of voxels (from 2000 to 4000) included as predictors (Baucom et al. 2012).

Another unrelated challenge for fostering the real-world applications of MBD of affective states is the high-cost and limited mobility of the so far most accurate imaging measurements (such as those acquired using fMRI or MEG) (Haynes and Rees 2006). In this context, functional Near Infrared Spectroscopy (fNIRS) emerges as an attractive neuroimaging technique for MBD-based affective discrimination. This method is based on low-energy light detectors and transmitters for measuring the light absorption through the cortical surface (Doi, Nishitani and Shinohara 2013). Thus, it makes possible to investigate local changes in the concentrations of oxy and deoxyhemoglobin in response to functional brain activity (Villringer et al. 1993), similarly to the widely known BOLD effect (Steinbrink et al. 2006). Furthermore, fNIRS has been considered a promising tool due to its good tradeoff between its spatial and temporal resolutions. It also presents low susceptibility to instrumental and biological noise when compared to EEG, and it is also less costly than fMRI (Doi, Nishitani and Shinohara 2013). These advantages introduce fNIRS as an attractive tool for portable MBD applications, including automatic diagnosis (Hock et al. 1996; Takizawa et al. 2014) and neurofeedback systems (Naseer and Hong 2015).

In light of the above-mentioned technical challenges for MBD, we aimed to evaluate whether fNIRS signals from frontal and occipital areas provide enough information to the inter-subject classification of affective states. To achieve this goal, we collected two independent databases with passive and active affective elicitation tasks. The passive elicitation condition was based on a set of International Affective Pictures System (IAPS) images that were thought to induce positive, negative or neutral affective states (Lang, Bradley and Cuthbert 1999). Critically, those images were selected in order to balance the arousal dimension for the different valences. In

the active elicitation condition, participants were instructed to imagine personal situations with positive, negative or neutral affective contexts. We expected decoding performances higher than chance level when comparing two different affective states in each elicitation condition. Also, based on the affective-workspace model (Edelman 1989; Barrett and Bliss-Moreau 2009), we expected the relevant information used in classification to come from nodes of neural networks mainly comprising frontal and occipital areas (Lindquist et al. 2012, 2015).

2. METHODS

2.1. Participants

Forty-nine healthy participants (25 female, 24.65±3.23 years) participated in the experiment. Inclusion criteria were no previous (self-reported) diagnosis of neurological (ICD-10: G00-G99) and/or psychiatric disorders (ICD-10: F00-F99), and normal or corrected-to-normal vision. All participants were attending college or graduated and provided written informed consent to participate in this study.

Two years later, a second (and independent) sample of twenty healthy participants (10 female, 24.00±3.92 years) was collected by a different researcher, and in a separate laboratory (at the same university). The inclusion criteria and the experimental protocol were equal to the first sample. For both cases, ethical approval was obtained from the local Ethics Committee, and no payment was provided to the participants, according to the national rules.

2.2. Functional NIRS acquisition

fNIRS measurements were conducted using the NIRScout System (NIRx Medical Technologies, LLC. Los Angeles, California) using an array of optodes (11 light sources/emitters and 11 detectors) covering the prefrontal, temporal and occipital areas. Optodes were arranged in an elastic band, with nine sources and nine detectors positioned over the frontotemporal regions and three sources and three detectors over the occipital region. Four positions of the International 10–20 System were adopted as reference points during the setup: sensors 1 and 9 were positioned approximately over the T7 and T8 locations, respectively, while the Fpz and Oz positions were in the center of channels 5-5 and 11-11, respectively, as shown in Figure 1A. The source-receptor distance was 30 mm for adjacent optodes, and the used wavelengths were 760 and 850 nm. Signals obtained from the thirty-two channels were measured with a sampling rate of 5.2083 Hz using the NIRStar 14.0 software (NIRx Medical Technologies, LLC. Los Angeles, California).

2.3. Experimental protocol

During the test, participants sat in a padded chair with armrest, positioned 1-meter distance in front of a monitor. They were asked to remain relaxed, with hands within sight resting on the armrests or the table. They were also requested to avoid eye movements, as well as any body movement. The recording room remained dark during registration and subject used earplugs.

Each subject completed an eleven-point Likert mood scale immediately before and after the session, to evaluate the possible influence of mood on our results. This test quantifies her/his agitation, strength, confusion, agility, apathy, satisfaction, worry, perspicacy, stress, attention, capacity, happiness, hostility, interest, and introspection (Stern 1997).

2.3.1. Block of passive elicitation

For the block of passive elicitation of affect, we used images available on the international affective picture system (IAPS) catalog (Lang, Bradley and Cuthbert 1999). First, images were filtered according to their average values of arousal and then were ranked according to their valence values. We selected the 30 pictures with higher average values of valence, the 30 with lowest values and 60 of intermediate values, as follows:

- Positive pictures (Valence = 7.884±0.220; Arousal = 5.036±0.448): 1811, 2057, 2080, 2209, 5210, 5830, 7200, 2040, 2058, 2091, 2340, 5700, 5833, 7330, 1440, 2045, 2070, 2150, 2347, 5825, 5910, 7502, 1710, 2050, 2071, 2165, 2550, 5829, 5982, 8420;
- Negative pictures (V = 2.007±0.183; A = 5.549±0.339): 2375.1, 3101, 3261, 9181, 9322, 9560, 2703, 3180, 3301, 9185, 9326, 9571, 2095, 2800, 3191, 3350, 9253, 9332, 2205, 3016, 3225, 9040, 9300, 9421, 2345.1, 3062, 3230, 9140, 9301, 9433;
- Neutral pictures (V = 5.234±0.060; A = 3.770±0.813): 2122, 2514, 5520, 7019, 7182, 7550, 2191, 2635, 5531, 7021, 7207, 7632, 2211, 2702, 5532, 7043, 7242, 7830, 1122, 2308, 2745.1, 5533, 7052, 7248, 8065, 1350, 2377, 2850, 5740, 7053, 7249, 1616, 2381, 2870, 5920, 7058, 7365, 1675, 2385, 2880, 6910, 7062, 7497, 1820, 2487, 5395, 7001, 7080, 7500, 1908, 2495, 5471, 7014, 7090, 7506, 2102, 2499, 5510, 7017, 7100.

This block consisted of twenty trials (5 for positive stimuli, 10 for neutral and 5 for negative). For the first 2 seconds of each trial, a white cross was presented in the center of a blank screen. During the next 30 seconds, a new figure was displayed every 5 seconds, totalizing six randomly selected figures per trial corresponding to the desired affective class. At the end of the trial, a new screen was presented asking the participant to assign a score from 1 to 9 for the subjective valence (1 – extremely negative valence; 9 – highly positive valence) and subjective arousal (1 – lower arousal; 9 – higher arousal) experiences. After this, a blank screen appears for a random duration between 2-4 seconds and participants were instructed to blink and/or move in this period but not in the other phases.

2.3.2. Block of active elicitation

The block of active elicitation of affect (affective imagination) consisted of twenty trials (5 trials for positive affect, 5 for negative affect and 10 for resting with eyes open, also called as neutral affect). Each trial started with a baseline period of a blank screen with a white cross in the center. After 2 seconds, the instruction (representing the desired emotion) appeared to the left of the display, remaining on the screen for 2 s (Figure 1B). The instruction consisted of either a green arrow pointing up (positive affect), a red arrow pointing downward (negative affect) or a blue circle (neutral affect).

For 30 seconds after the instruction was presented, the screen remained unchanged, corresponding to the participant's affective imagination period. At the end of the trial, a new screen was presented asking the participant to assign a score from 1 to 9 for the subjective valence (1 – extremely negative valence; 9 – highly positive valence) and subjective arousal (1 – lower arousal; 9 – higher arousal) experiences. After this, a blank screen appears for a random duration between 2 to 4 seconds and participants were instructed to blink and/or move in this period.

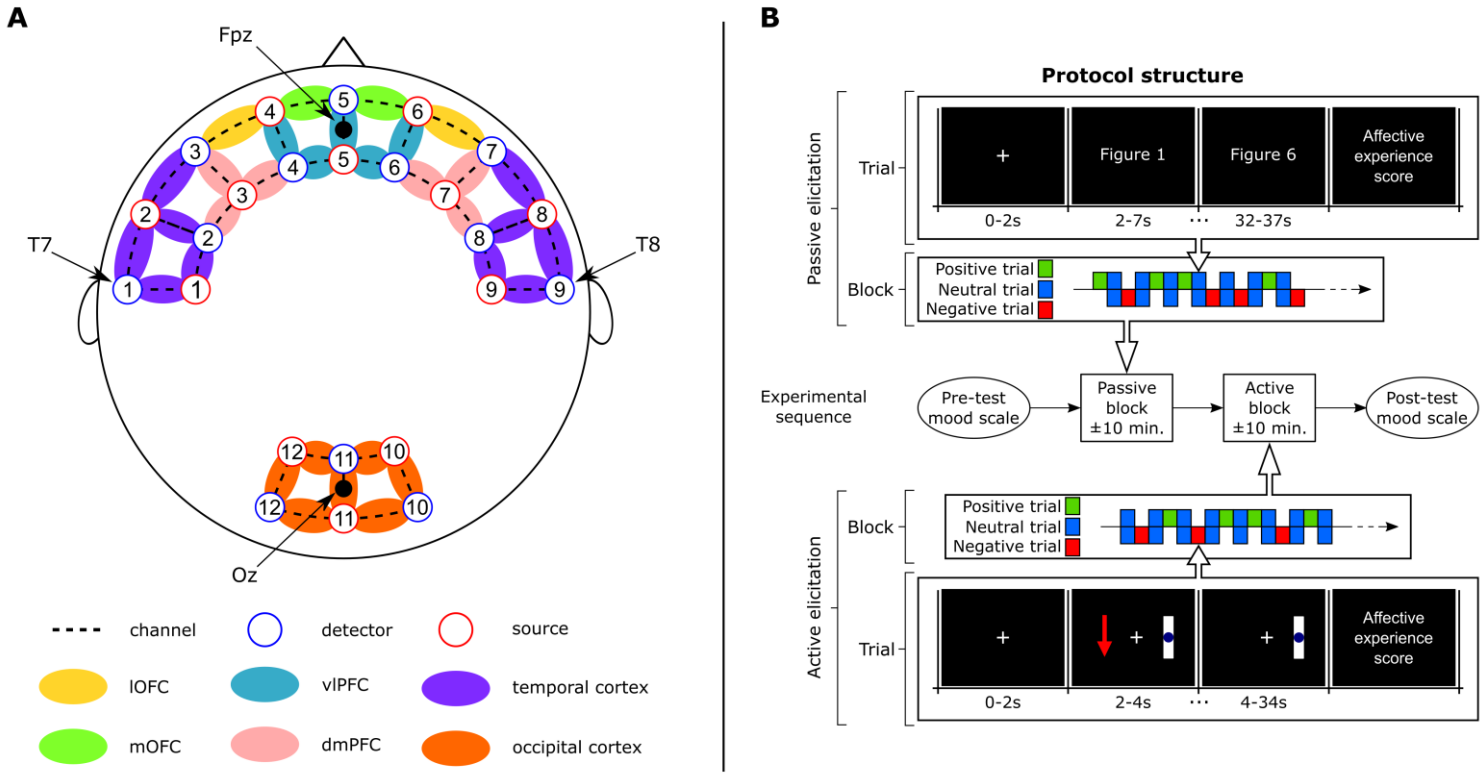

Figure 1 – (A) Channel configuration. Red circles represent sources; blue circles represent the detectors and dotted lines the channels. (B) Visual stimuli order in passive and active elicitation trials. The order of trials into the block is random, but always alternating neutral trials with positive and negative affect trials.

2.4. Data analysis

2.4.1. Preprocessing

Preprocessing was performed using Matlab (Mathworks, MA, USA) with the nirsLAB v2014.12 toolbox (NIRx Medical Technologies, LLC. Los Angeles, California). Each participant's raw data were digitally band-pass filtered by a linear-phase FIR filter (0.01-0.2 Hz). Then each wavelength was detrended by their respective whole length record (without segmentation), and the concentration curves of oxyhemoglobin and deoxyhemoglobin were calculated by the Beer-Lambert law (differential pathlength factor (DPF) set to 7.25 and 6.38, respectively) (Essenpreis et al. 1993). Each concentration curve was then segmented into the 30 s of interest of each trial, for all studied conditions.

The mean concentration of oxyhemoglobin and deoxyhemoglobin for each segment was calculated for each channel using the average of moving 2s-window means with 50% overlap. Thus, each subject's database was composed of 64 features of average concentration (32 channels x 2 chromophores) in 20 experimental conditions (10 neutral trials + 5 positive trials + 5 trials), for both blocks.

2.4.2. Across databases classification

To investigate inter-subject affective decoding, we used a cross-validation approach (Pereira, Botvinick and Mitchell 2009). The first experimental sample (49 participants) was used for the model training, while the second sample (20 participants) was used to evaluate the prediction accuracy. This approach was applied to evaluate affective conditions in pairs, following "Positive versus Neutral", "Negative versus Neutral" and "Positive versus Negative" combinations, for blocks of passive and active elicitation separately.

A Linear Discriminant Analysis (LDA) method was applied in these analyses using the BCILAB toolbox (Kothe and Makeig 2013) implementation with default parameters. This method performs a prediction based on a linear combination of all features with its respective weights (Scholkopft and Mullert 1999). Another characteristic of LDA is its focus on extracting all discriminative information available in the sample, which might provide critical information to decode neural phenomena (Sato et al. 2009).

To evaluate the prediction performance in each comparison, we calculated the decoding accuracy as (trials correctly classified from first class + trials correctly classified from second class)/2. This approach avoids the potential unbalance of sample size in each comparison, and this random result should be 50%.

2.4.3. Feature selection

As previously mentioned, each LDA model creates a linear combination of all features with its respective weights. Thus, the higher the absolute weight, the more important the feature is (Scholkopft and Mullert 1999). In order to reduce the number of features used during classification, relevant features were selected using the weights of each feature. For this, we ranked the absolute weights assigned by the classifier model in each training set. We selected the best 5–100% (with steps of 5%) features with highest weights, separately, and submitted each new features sample to re-train the classifier and test it with the second database. This configuration was performed for feature sets containing oxyhemoglobin and deoxyhemoglobin simultaneously.

2.4.4. Statistical analysis

First, the subject scores of valence and arousal of the second database were evaluated by a two sample t-test. The comparisons used the mean values of valence and arousal assigned during positive, negative and neutral trials. This procedure was repeated for both elicitation blocks. The p-values were Bonferroni corrected for multiple comparisons (2 subjective measures x 3 comparisons of conditions x 2 elicitation blocks).

The significance of classification accuracy of each comparison was evaluated by a one sample t-test, comparing the prediction performance from all participants in this comparison against chance level (50%). Each result was independently adjusted by using the Bonferroni correction for 120 multiple comparisons (20 different percentages of features x 3 comparisons of conditions x 2 elicitation blocks).

To rule out potential confounders due to the experimental procedure itself on the results, we calculated the Spearman correlation between decoding accuracies and the variation of each measured mood variables in the Likert scales applied immediately before and after the experiment. The p-values were also Bonferroni corrected for multiple comparisons (16 mood variables x 120 accuracy variables). Finally, putative effects of gender were examined by comparing the decoding accuracies separately for men and women. A t-test for two independent samples was performed (Bonferroni corrected for the 120 tests).

3. RESULTS

3.1. Subjective scores of Arousal and Valence

Table 1 presents the mean subjective scores of arousal and valence relative to the positive, negative and neutral conditions, during both active and passive elicitation. Moreover, the corrected p-values of each comparison are also presented.

Table 1 – In the first half of the table, mean ± standard-deviation of subjective scores (ranging from 1 to 9) assigned by each subject of the second database during negative, neutral and positive trials. In the second half, p-values of the comparison between the scores assigned for each trial type.

Mean subjective scores		
	Valence	**Arousal**
Passive elicitation		
Negative	2.52 ± 1.38	6.05 ± 1.34
Neutral	5.11 ± 0.73	2.71 ± 1.29
Positive	6.79 ± 0.92	4.82 ± 1.60
Active elicitation		
Negative	2.57 ± 0.97	5.85 ± 1.71
Neutral	4.84 ± 0.24	2.09 ± 1.39
Positive	7.20 ± 0.77	5.67 ± 1.75
Statistical comparisons (p-values)		
	Valence	**Arousal**
Passive elicitation		
Positive vs. Negative	*<0.01*	0.14
Positive vs. Neutral	*0.02*	1.00
Negative vs. Neutral	*<0.01*	*<0.01*
Active elicitation		
Positive vs. Negative	*<0.01*	1.00
Positive vs. Neutral	*<0.01*	*<0.01*
Negative vs. Neutral	*<0.01*	*<0.01*

Valence scores presented significant differences for all the comparisons. It is relevant, however, that no significant difference of the arousal scores was found during

Positive vs. Negative comparisons, as well as during Positive vs. Neutral comparison during the passive block.

3.2. Classification results

LDA accuracy boxplots are presented in Figure 2. Classification accuracy for passive elicitation significantly exceeded chance level in “Positive vs. Negative” comparisons, with highest result as 64.50±12.44% ($p<0.001$) using 20% of features (12 channels), and in “Negative vs. Neutral” comparisons (67.75±14.45%, $p<0.01$, 40% of features – 25 channels). No significant decoding effects were found for “Positive vs. Neutral” comparisons.

During active elicitation comparisons, “Positive vs. Neutral” accuracies were different to chance, where the highest result was 71.00±17.95% ($p<0.001$) using only 5% of features (3 channels). No significant decoding effect was found for “Positive vs. Negative” and “Negative vs. Neutral” comparisons.

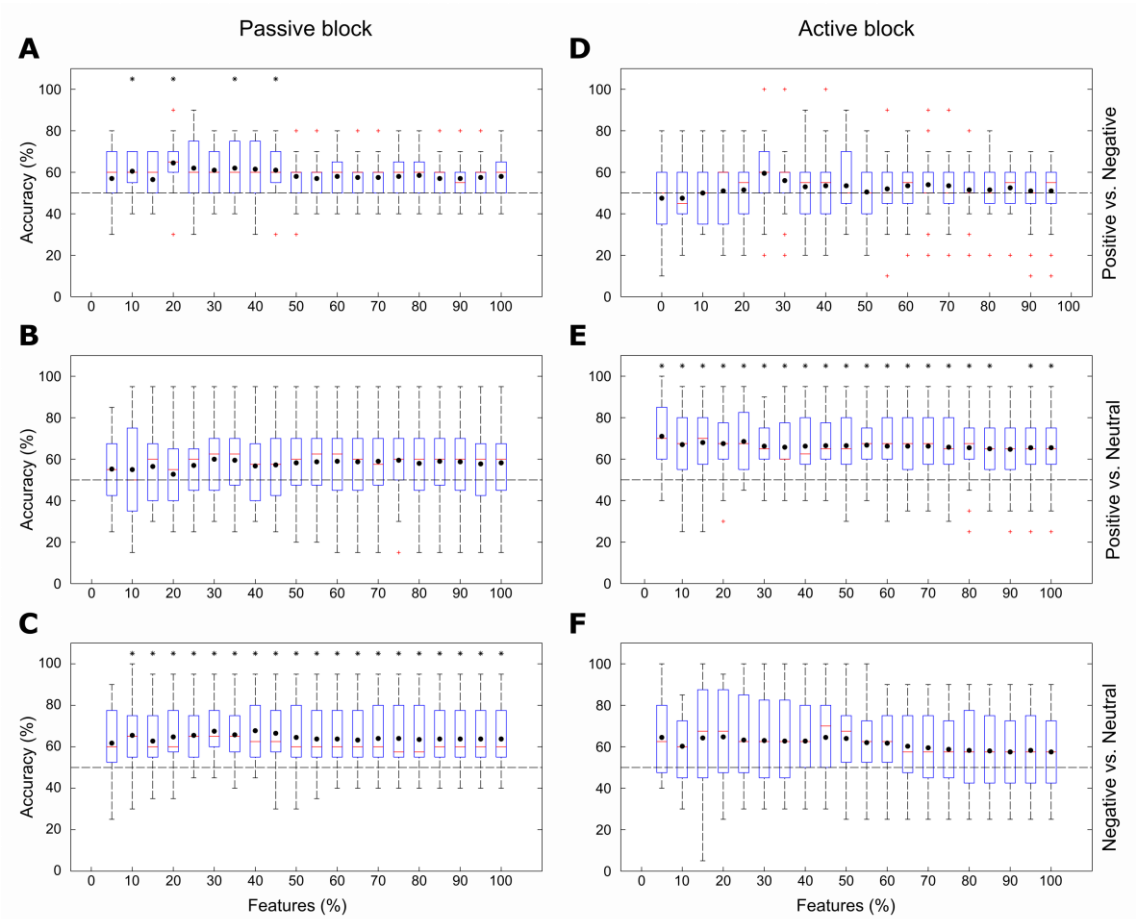

Figure 2 - Box plots showing the results using the LDA classifier. (A) presents the “Positive vs. Negative”, (B) the “Positive vs. Neutral”, and (C) the “Negative vs. Neutral” comparisons for passive elicitation block, while (D-F) follows the same order for active elicitation block. Black dots present the means, red lines the medians, red crosses the outliers and black asterisks the statistical difference for chance level ($p<0.05$).

No gender effect was observed for the valence prediction. Also, no significant correlations were found between mood variations and the prediction scores.

3.3. Relevant channels

In Figure 3 we present the channels selected in each of the highest significant results previously reported, and its respective weights assigned the classifier. To this end, the location of the channels follows the order shown in Figure 1A.

Among the Positive vs. Negative classification during the passive elicitation block, the highest accuracy was achieved using 12 features. Nine of these features correspond to information about the deoxyhemoglobin concentration above the medial orbitofrontal cortex (mOFC, one channel), the ventrolateral prefrontal cortex (vlPFC, three), the dorsomedial prefrontal cortex (dmPFC, one), the lateral orbitofrontal cortex (lOFC, one) and the occipital cortex (three). Complementarily, three features included information about the oxyhemoglobin concentration above the vlPFC (two channels) and the occipital cortex (one).

For the Negative vs. Neutral classification during the passive elicitation, 25 channels were carried relevant information: two mOFC channels, four vlPFC, four dmPFC, two lOFC, two temporal and three occipital channels for the deoxyhemoglobin concentration. For the oxyhemoglobin, two mOFC, one lOFC, one vlPFC, one dmPFC and three occipital channels were considered relevant.

Finally, the Positive vs. Neutral decoding during the active elicitation block used only three channels to achieve the highest performance: two vlPFC and one lOFC. These channels provided information about the deoxyhemoglobin concentration.

4. DISCUSSION

4.1. Across databases classification

First, it is essential to highlight the absence of statistical significance in the subjective arousal score during the “Positive vs. Negative” comparisons. In addition to the significant difference in the valence scores, it suggests that any classification result in these comparisons is exclusively related to valence differences between positive and negative affect. Further, during both “Positive vs. Neutral” and “Negative vs. Neutral” comparisons, statistical differences were found for valence and arousal scores. It is expected that Positive and Negative present differences in arousal and valence when compared to the Neutral affect. Some studies suggest valence and arousal not as two entirely separable dimensions, but as a single V-shaped dimension (Kuppens et al. 2013). This conceptual caveat is also experimentally founded, since it is not possible to separate the two aspects in most of the emotional stimuli systems (for example, the International Affective Pictures System) (Lang, Bradley and Cuthbert 2005), with valence accompanied by arousal changes and vice versa (Barrett and Bliss-Moreau 2009). However, considering clinical applications in affective disorders, for example, it is not relevant to an MBD to differentiate between valence and arousal. Otherwise, the primary focus is to train the user to be able to learn the regulation of his/her

neurophysiological activity, offering the possibility to normalize the activity level and reduce the symptom severity (Fovet, Jardri and Linden 2015).

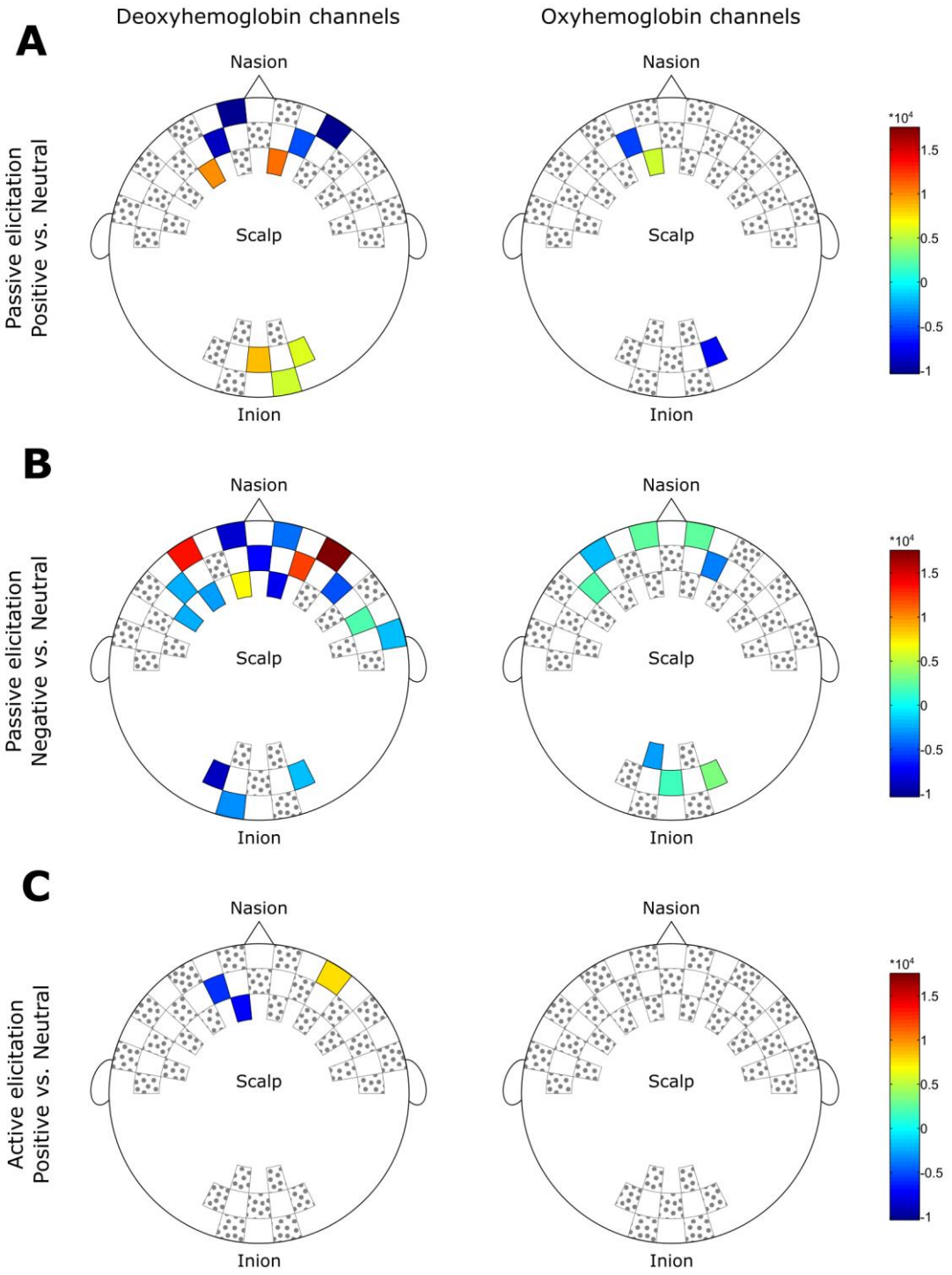

Figure 3 - Weights assigned by the LDA classifier for each feature during the (A) “Positive vs. Negative” comparisons using 20% of features and (B) “Negative vs. Neutral” comparisons using 40% of features. In (C), the weights for each feature during the “Positive vs. Neutral” comparisons using 5% of features. Hottest colors indicate positive weights while cooler colors indicate negative weights. Channels filled with gray dots were not used during the test.

Given this, we have demonstrated that it is possible to detect distinct patterns of hemodynamic activity generated by different affective valence elicitation tasks. Even more, we also evidenced that it is feasible to decode affective states based on a model trained with an independent dataset. The decoding of passive elicitation, for example, achieved statistical significance exclusively during the “Positive vs. Negative” and “Negative vs. Neutral” comparison. This finding corroborates with previous inter-subject studies which also used the IAPS database to successful decode affective states using fMRI data (Baucom et al. 2012). The absence of significant difference for the “Positive vs. Neutral” condition is unexpected, considering that the IAPS database is strictly standardized to produce these affective states (Lang, Bradley and Cuthbert 1999, 2005). The most likely explanation is the lack of subcortical measures in our test as well as our limited number of cortical channels. Regions such as the anterior cingulate, fusiform gyrus, superior temporal lobe, lentiform nucleus, parahippocampus, and precuneus are listed as informative voxel location clusters across participants in previous experiments (Baucom et al. 2012) but were not recorded with our setup due to our experimental limitations.

The decoding of active elicitation, otherwise, presented significant accuracy for the “Positive vs. Neutral” condition exclusively. It is a relevant result considering possible neurofeedback applications in psychiatry since it requires a self-regulation of neural activity related to a specific affective task (Birbaumer, Ruiz and Sitaram 2013, Kim and Birbaumer 2014). In this case, the most interesting combination of affective states is the “Positive vs. Neutral”, since patients might present increased levels of negative emotions and decreased levels of positive emotions (Gross and Munhoz 1995). Even more, mean accuracy was slightly over the 70% threshold suggested by the brain-computer interface and neurofeedback communities as sufficient to perform device control and communication (Tai and Chau 2009; McFarland et al. 2006). This way, we can infer that an affective neurofeedback trained with a fixed and independent training set is a possible resource for future psychiatric applications. It is also relevant that this result was reached using only three channels. Still considering future neurofeedback applications, the use of exclusively three channels means shorter setup procedure, lower instrumental and computational costs and the possibility of systems even more portable.

Finally, some participants described after the experiment that the negative affect induced by the pictures presented during the passive block was more intense than the positive one. Negative figures into the IAPS catalog are mainly related to death, malnutrition, sickness, poverty, and disgust (Lang, Bradley and Cuthbert 1999), which are more consensual than some content of the positive stimuli, such as babies, pets or beaches. Moreover, some studies suggest that processing negative stimuli are more demanding in the brain (Ito et al. 1998), which might generate more clear signals to our classifiers than the neural processing of positive stimuli. On the other hand, the participants also described that the positive imagery was more natural to achieve and more intense than the negative representation. In our daily life, we experience much more negative than positive emotions, and it is used to improve learning and regulate attention, for example (Greenberg 2004). However, it is fundamental to the mental health to compensate and regulate these negative emotions (Gross 1998). In this context, a common strategy is the use of affective imagery, which simulates and

recreate personal experiences (Pearson et al. 2013), such as involving family, friends, relationships. This expertise might be a possible explanation to the reported easiness and intensity of the positive affect imagery and, consequently, the significance of the results during the active block.

4.2. Relevant features and the neural networks of affect

For both passive elicitation results, relevant features included oxyhemoglobin and deoxyhemoglobin concentrations. This result is not entirely unexpected once each chromophore is expected to provide different information. Deoxyhemoglobin is thought to be an indicator of functional activation and commonly related to the fMRI BOLD signal (Song, Huettel and McCarthy 2006, Steinbrink et al. 2006). On the other hand, a recent study correlated the oxyhemoglobin curves with the EEG band power variation in some cortical regions (Pfurtscheller et al. 2012). In the same way, other decoding experiments also reported both hemoglobin concentrations as relevant for classification (Tai and Chau 2009; McFarland et al. 2006). These results suggest a non-redundancy of these measures.

In accordance with two meta-analyses of neuroimaging data in affective tasks, we found that occipital areas signals were among the relevant features for classification (Lindquist et al. 2012, 2015). Complementarily, fNIRS studies also described activation of the occipital cortex during affective stimulation (Alpers et al. 2004, Herrmann et al. 2008). Minati et al. (2009), for example, found increased amplitude response to positive and negative affects relative to neutral pictures of the same database we have used (IAPS).

Moreover, it is interesting to note that the classifier assigned high absolute weights for a considerable amount of channels from frontotemporal regions. In general, a classical effect during affective tasks is the frontotemporal lateralization during the experience and regulation of emotional responses (Ahern and Schwartz 1985; Ochsner and Gross, 2005; Balconi, Bortolotti and Gonzaga 2011). Although the standard effect includes positive affective valence more directly processed by the left hemisphere and negative affective valence by the right hemisphere (Ahern and Schwartz 1985; Everhart et al. 2003; Balconi and Mazza, 2010), our highest results were classifying positive or negative affects versus neutral trials. A possible explanation of this is that our data agree with the affective workspace hypothesis in which activity patterns of the same core neural network could implement both positive and negative affective responses (Edelman 1989; Barrett and Bliss-Moreau 2009).

Considering channels from central regions, the right vlPFC area has been reported as a component of two pathways responsible to the reappraisal of aversive images, involving the nucleus accumbens and ventral amygdala (Wager et al. 2008), while the lOFC is considered a heteromodal association area including internal and external sensory representations (Barrett and Bliss-Moreau 2009). Likewise, the dmPFC region, which was also relevant for classification in both trials, is a fundamental default mode network node and has been proposed as a core region for the affective neural workspace (Barrett et al. 2007; Lindquist et al. 2012). Remarkably, Chikazoe et al. (2014) found populations of neurons responding to both positive and negative affective stimuli in several brain regions, including the mOFC and lOFC, which could

also be related to the relevance of these areas in the classification between affectively charged and neutral trials.

4.3. Limitations and future perspectives

Despite the controls we implemented in our experimental design, several limitations should be recognized. First, although fNIRS mainly measures the near-infrared light absorption into the cortical surface, it also encompasses peripheral responses to task stimuli, such changes in superficial tissue blood flow, blood pressure, heart rate and the aerobic process of energy consumption related to muscle contraction (Doi, Nishitani and Shinohara 2013). Consequently, some authors refer to fNIRS applications for control of devices such as corporeal machine interfaces (Tai and Chau 2009). Further study of the issue is still required and should involve psychophysiological records associated with neuroimaging data.

However, even with these limitations, we reinforce the advantages of fNIRS compared to traditional MBD systems, in particular, its benefit-cost and the proper relation of its spatial and temporal resolutions (Doi, Nishitani and Shinohara 2013). These aspects allow a fast replication and generalization of our inter-subject decoding results to different tasks and databases such as motor imagery, mental arithmetic and face recognition (Tong and Pratte 2012), or during real context experiments outside the lab (Falk et al. 2011).

Finally, our experiment focused on a primary processing step, classifying fNIRS data using mean changes in oxy and deoxyhemoglobin concentrations as discriminant features. This approach is indeed the simplest and quite a usual way to analyze NIRS data (Coyle, Ward and Markham 2007; Sitaram et al. 2007). In comparison, recent studies identified that combining the mean hemoglobin concentration with other temporal and time-frequency features improves the decoding accuracies reaching values close to 90% in within-subject decoding (Tai and Chau 2009). Therefore, future studies should also evaluate the effect of different feature extraction techniques to the inter-participants MBD.

5. CONCLUSION

Our results demonstrated that fNIRS is a suitable tool for inter-participants multivariate brain decoding, especially those involving the evaluation of affective states. The accuracy measures are significant and above the threshold desired for effective control of brain-computer interfaces and the relevant features are consistent with previous affective neuroscience findings. Consequently, future research might explore the use of inter-participants training sets in future fNIRS-based BCI or neurofeedback devices.

6. ACKNOWLEDGEMENT

This work was supported by UFABC, CAPES, and FAPESP (grant numbers 2015/17406-5 and 2017/05225-1).

References

Agrafioti, F., Hatzinakos, D., Anderson, A.K., 2012. ECG pattern analysis for emotion detection. IEEE Trans. Affect. Comput., 3, 102-115.

Ahern, G.L., Schwartz, G.E., 1985. Differential lateralization for positive and negative emotion in the human brain: EEG spectral analysis. Neuropsychologia, 23, 745-755.

Alpers, G.W., Herrmann, M.J., Pauli, P., Fallgatter, A.J., 2004. Emotional arousal and activation of the visual cortex: A near infrared spectroscopy analysis. J. Psychophysiol., 19, 106.

Balconi, M., Mazza, G., 2010. Lateralisation effect in comprehension of emotional facial expression: a comparison between EEG alpha band power and behavioural inhibition (BIS) and activation (BAS) systems. Laterality, 15, 361-384.

Balconi, M., Bortolotti, A., Gonzaga, L., 2011. Emotional face recognition, EMG response, and medial prefrontal activity in empathic behaviour. Neurosci. Res., 71, 251-259.

Barrett, L.F., Mesquita, B., Ochsner, K.N., Gross, J.J., 2007. The experience of emotion. Annu. Rev. Psychol., 58, 373-403.

Barrett, L.F., Bliss-Moreau, E., 2009. Affect as a psychological primitive. Adv. Exp. Soc. Psychol., 41, 167-218.

Baucom, L.B., Wedell, D.H., Wang, J., Blitzer, D.N., Shinkareva, S.V., 2012. Decoding the neural representation of affective states. Neuroimage, 59, 718-727.

Birbaumer, N., Ruiz, S., Sitaram, R., 2013. Learned regulation of brain metabolism Trends Cogn. Sci., 17, 295-302

Bozhkov, L., Petia, G., 2014. Classification models of emotional biosignals evoked while viewing affective pictures, in: International Conference on Simulation and Modeling Methodologies, Technologies and Applications. IEEE, pp. 601-606.

Chikazoe, J., Lee, D.H., Kriegeskorte, N., Anderson, A.K., 2014. Population coding of affect across stimuli, modalities and individuals. Nat. Neurosci., 17, 1114-1122.

Coyle, S.M., Ward, T.E., Markham, C.M., 2007. Brain–computer interface using a simplified functional near-infrared spectroscopy system. J. Neural Eng., 4, 219.

Doi, H., Nishitani, S., Shinohara, K., 2013. NIRS as a tool for assaying emotional function in the prefrontal cortex. Front. Hum. Neurosci., 7, 770.

Edelman, G.M., 1989. The remembered present: A biological theory of consciousness, first ed. Basic Books, New York.

Essenpreis, M., Elwell, C.E., Cope, M., Van der Zee, P., Arridge, S.R., Delpy, D.T., 1993. Spectral dependence of temporal point spread functions in human tissues. Appl. Opt., 32, 418-425.

Everhart, D.E., Carpenter, M.D., Carmona, J.E., Ethridge, A.J., Demaree, H.A., 2003. Adult sex-related P300 differences during the perception of emotional prosody and facial affect. Psychophysiology, 40, S39-S39.

Falk, T.H., Guirgis, M., Power, S., Chau, T.T., 2011. Taking NIRS-BCIs outside the lab: towards achieving robustness against environment noise. IEEE Trans. Neural Syst. Rehabil. Eng., 19, 136-146.

Fovet, T., Jardri, R., Linden, D., 2015. Current issues in the use of fMRI-based neurofeedback to relieve psychiatric symptoms. Curr. Pharm. Des., 21, 3384-3394.

Georgieva, O., Milanov, S., Georgieva, P., Santos, I.M., Pereira, A.T., Silva, C.F., 2015. Learning to decode human emotions from event-related potentials. Neural Comput. Appl., 26, 573-580.

Greenberg, L.S., 2004. Emotion–focused therapy. Clin. Psychol. Psychother., 11, 3-16.

Gross, J.J., Muñoz, R.F., 1995. Emotion regulation and mental health. Clin. Psychol., 2, 151-164.

Gross, J.J., 1998. The emerging field of emotion regulation: An integrative review. Rev. Gen. Psychol., *2*, 271-299.

Haynes, J.D., Rees, G., 2006. Decoding mental states from brain activity in humans. Nat. Rev. Neurosci., 7, 523-534.

Herrmann, M.J., Huter, T., Plichta, M.M., Ehlis, A.C., Alpers, G.W., Mühlberger, A., Fallgatter, A.J., 2008. Enhancement of activity of the primary visual cortex during processing of emotional stimuli as measured with event-related functional near-infrared spectroscopy and event-related potentials. Hum. Brain Mapp., 29, 28-35.

Hock, C., Villringer, K., Müller-Spahn, F., Hofmann, M., Schuh-Hofer, S., Heekeren, H., Wenzel, R., Dirnagl, U. Villringer, A., 1996. Near Infrared Spectroscopy in the Diagnosis of Alzheimer's Diseasea. Ann. N. Y. Acad. Sci., 777, 22-29.

Ito, T.A., Larsen, J.T., Smith, N.K., Cacioppo, J.T., 1998. Negative information weighs more heavily on the brain: the negativity bias in evaluative categorizations. J. Pers. Soc. Psychol., 75, 887-900.

Kim, S., Birbaumer, N., 2014. Real-time functional MRI neurofeedback: a tool for psychiatry Curr. Opin. Psychiatry, 27, 332-336

Kothe, C.A., Makeig, S., 2013. BCILAB: a platform for brain–computer interface development. J. Neural Eng., 10, 056014.

Kubilius, J., Baeck, A., Wagemans, J., de Beeck, H.P.O., 2015. Brain-decoding fMRI reveals how wholes relate to the sum of parts. Cortex, 72, 5-14.

Kuppens, P., Tuerlinckx, F., Russell, J.A., Barrett, L.F., 2013 The relation between valence and arousal in subjective experience. Psychol. Bull., 139, 917-940.

Lang, P.J., Bradley, M.M., Cuthbert, B.N., 1999. International affective picture system (IAPS): Technical manual and affective ratings.

Lang, P.J., Bradley, M.M., Cuthbert, B.N., 2005. International Affective Picture System (IAPS): Affective Ratings of Pictures and Instruction Manual. Technical Report A-8, NIMH Center for the Study of Emotion & Attention.

Lawrence, E.J., Su, L., Barker, G.J., Medford, N., Dalton, J., Williams, S.C., Birbaumer, N., Veit, R., Ranganatha, S., Bodurka, J., Brammer, M., Giampietro, V., David, A.S., 2014. Self-regulation of the anterior insula: Reinforcement learning using real-time fMRI neurofeedback. Neuroimage, 88, 113-124.

Lindquist, K.A., Wager, T.D., Kober, H., Bliss-Moreau, E., Barrett, L.F., 2012. The brain basis of emotion: a meta-analytic review. Behav. Brain Sci., 35, 121-143.

Lindquist, K.A., Satpute, A.B., Wager, T.D., Weber, J., Barrett, L.F., 2015. The brain basis of positive and negative affect: evidence from a meta-analysis of the human neuroimaging literature. Cereb. Cortex, bhv001.

Mariooryad, S., Carlos B., 2013. Exploring cross-modality affective reactions for audiovisual emotion recognition. IEEE Trans. Affect. Comput., 4, 183-196.

McFarland, D.J., Anderson, C.W., Muller, K.R., Schlogl, A., Krusienski, D.J., 2006. BCI meeting 2005-workshop on BCI signal processing: feature extraction and translation. IEEE Trans. Neural Syst. Rehabil. Eng., 14, 135-138.

Minati, L., Jones, C.L., Gray, M.A., Medford, N., Harrison, N.A., Critchley, H.D., 2009. Emotional modulation of visual cortex activity: a functional near-infrared spectroscopy study. Neuroreport, 20, 1344.

Naseer, N., Hong, K.S., 2015. fNIRS-based brain-computer interfaces: a review. Front. Hum. Neurosci., 9, 3.

Ochsner, K.N., Gross, J.J., 2005. The cognitive control of emotion. Trends Cogn. Sci., 9, 242-249.

Pearson, D.G., Deeprose, C., Wallace-Hadrill, S.M., Heyes, S.B., Holmes, E.A., 2013. Assessing mental imagery in clinical psychology: A review of imagery measures and a guiding framework. Clin. Psychol. Rev., 33, 1-23.

Pereira, F., Mitchell, T., Botvinick, M., 2009. Machine learning classifiers and fMRI: a tutorial overview. Neuroimage, 45, S199-S209.

Pfurtscheller, G., Daly, I., Bauernfeind, G., Müller-Putz, G.R., 2012. Coupling between intrinsic prefrontal HbO2 and central EEG beta power oscillations in the resting brain. PLoS One, 7, e43640.

Ramaraju, S., Izzidien, A., Roula, M.A., 2015. The detection and classification of the mental state elicited by humor from EEG patterns, in: International Conference of the IEEE. Engineering in Medicine and Biology Society. IEEE, pp. 1472-1475.

Sato, J.R., Fujita, A., Thomaz, C.E., Martin, M.D.G.M., Mourão-Miranda, J., Brammer, M.J., Junior, E.A., 2009. Evaluating SVM and MLDA in the extraction of discriminant regions for mental state prediction. NeuroImage, 46, 105-114.

Scholkopft, B., Mullert, K.R., 1999. Fisher discriminant analysis with kernels, in: Hu, Y., Larsen, J., Wilson, E., Douglas, S. (Eds.), Neural networks for signal processing IX. IEEE, New York, pp 41-48.

Sitaram, R., Zhang, H., Guan, C., Thulasidas, M., Hoshi, Y., Ishikawa, A., Shimizu, K., Birbaumer, N., 2007. Temporal classification of multichannel near-infrared spectroscopy signals of motor imagery for developing a brain–computer interface. NeuroImage, 34, 1416-1427.

Song, A.W., Huettel, S.A., McCarthy, G., 2006. Functional neuroimaging: Basic principles of functional MRI, in: Cabeza, R., Kingstone, A. (Eds.), Handbook of functional neuroimaging of cognition. MIT Press, Cambridge, pp. 22-52.

Steinbrink, J., Villringer, A., Kempf, F., Haux, D., Boden, S., Obrig, H., 2006. Illuminating the BOLD signal: combined fMRI–fNIRS studies. Magn. Reson. Imaging, 24, 495-505

Stern, R.A., 1997. VAMS: Visual analog mood scales: Professional manual. Odessa, FL: Psychological Assessment Resources

Tai, K., Chau, T., 2009. Single-trial classification of NIRS signals during emotional induction tasks: towards a corporeal machine interface. J. Neuroeng. Rehabil., 6, 39.

Takizawa, R., Fukuda, M., Kawasaki, S., Kasai, K., Mimura, M., Pu, S., Noda, T., Niwa, S., Okazaki, I, Joint Project for Psychiatric Application of Near-Infrared Spectroscopy (JPSY-NIRS) Group, 2014. Neuroimaging-aided differential diagnosis of the depressive state. Neuroimage, 85, 498-507.

Tan, J.W., Andrade, A.O., Li, H., Walter, S., Hrabal, D., Rukavina, S., Limbrecht-Ecklundt, K., Hoffman, H., Traue, H.C., 2016. Recognition of Intensive Valence and Arousal Affective States via Facial Electromyographic Activity in Young and Senior Adults. PloS one, 11, e0146691.

Tong, F., Pratte, M.S., 2012. Decoding patterns of human brain activity. Annu. Rev. Psychol., 63, 483-509.

Villringer, A., Planck, J., Hock, C., Schleinkofer, L., Dirnagl, U., 1993. Near infrared spectroscopy (NIRS): a new tool to study hemodynamic changes during activation of brain function in human adults. Neurosci. Lett., 154, 101-104.

Wager, T.D., Davidson, M.L., Hughes, B.L., Lindquist, M.A., Ochsner, K.N., 2008. Prefrontal-subcortical pathways mediating successful emotion regulation. Neuron, 59, 1037-1050.