Analysis of the Impact of the Presence of Physical Pain in fNIRS-based BCI Systems

A. Subramanian[1], F. Shamsi[2] and L. Najafizadeh[1]

1. Department of Electrical and Computer Engineering, Rutgers University, Piscataway, NJ 08854, USA
2. Department of Psychology, Northeastern University, Boston, MA 02115, USA
{ashwini.subramanian, laleh.najafizadeh}@rutgers.edu, f.shamsi@northeastern.edu

***Abstract*— An important application of brain computer interface devices (BCIs) is in assistive systems for patients with motor and communication disabilities. Due to their condition, these patients may experience pain. However, how the presence of pain influences the operation of such BCIs has not been fully investigated. This paper studies the impact of the presence of acute pain on the classification accuracy of a BCI, which employs functional near infrared spectroscopy (fNIRS) for brain signal acquisition. Cortical signals are obtained in the presence and absence of an external pain stimulus, while participants perform two mental arithmetic tasks. Convolutional neural network (CNN) is used to classify the tasks. It is observed that when the classifier is trained on pain-free data and tested on data obtained in the presence of pain, the classification accuracy significantly drops. Next, multi-label classification is performed to simultaneously identify the presence of pain and classify the tasks, further demonstrating that the distinction of tasks in the presence of pain is challenging. Finally, to mitigate the impact of pain, it is proposed to train the model collectively on data obtained in the presence and the absence of pain. It is observed that using this approach significantly improves the classification accuracy. Our results suggest that it is critical to include data obtained in the presence of pain in the training process of the classification models, when designing BCIs in assistive systems for patients.**

I. Introduction

Non-invasive functional neuroimaging techniques provide unique opportunities for studying the functionality of the brain in realistic environments, as well as for the realization of assistive systems that rely on brain computer interfaces (BCIs). The role of a BCI in such systems is to create commands for controlling peripheral devices (e.g., a wheelchair or a robotic arm), from brain signals that correspond to user's intentions. For the realization of the brain signal acquisition module in BCIs, while electroencephalography (EEG) has been widely used [1], functional near infrared spectroscopy (fNIRS) has been gaining interests to be employed, due to its advantages such as being less sensitive to motion artifacts [2]-[5].

Pain is prevalent in a substantial number of patients with motor disabilities who are in need of assistive systems. The presence of pain could be independent of the underlying condition and last a long duration (chronic pain condition), or it could have an unpredictable onset with a relatively short duration (acute pain). Few studies have tried to identify biomarkers for pain [6]-[10] in an attempt to standardize pain measurement. However, till date, self-reporting is the commonly-used method for pain description, making it highly subjective. Pain is known to impact cortical activities, influencing cognitive functions [11]-[12]. As such, an important question that needs to be investigated would be: is the operation of BCIs compromised by the presence of pain? For example, how would the BCI perform if the patient does not experience pain while training the BCI, but later experiences pain while using the BCI? The limited existing research in this area suggests that EEG and fNIRS signals are influenced by the presence of pain [13]-[16], highlighting the need for finding models and features that are robust to sudden occurrences of pain.

In this paper, we investigate the impact of the presence of acute pain on the classification accuracy of a fNIRS-based BCI. We consider a one dimensional convolutional neural network (CNN) as the classifier for the BCI. In our previous work [16], we considered frequency-domain features from fNIRS signals with support vector machine (SVM) as the classifier. SVM required extensive manual crafting of features. As an alternative approach, here, we use CNN, as it learns features automatically from the input data, thereby, eliminating the need for manual feature extraction. Next, we perform a multi-label classification to identify the presence of pain and classify the performed task, simultaneously. Finally, we propose a training process that pools data obtained both in the presence and the absence of pain as a solution to mitigate the presence of pain for the operation of the BCI.

The paper is organized as follows. Experimental procedure, data preprocessing steps and classification methods are described in Section II. Results and discussions are presented in Section III, and the paper is concluded in Section IV.

II. Methods

II-A. Experimental procedure

All procedures were approved by the Rutgers' Institutional Review Board (IRB). Data collection from 3 healthy subjects were performed in the NeuroImaging Laboratory at Rutgers University. Details of the experiment have been previously described in [16].

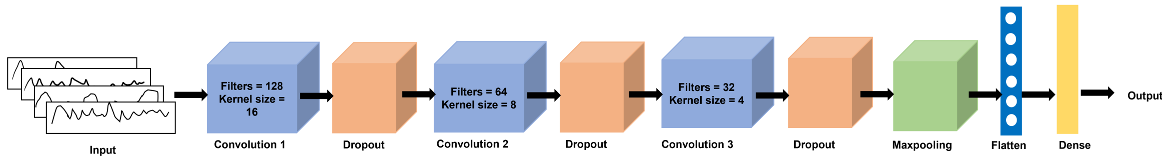

Figure 1. Architecture of the CNN

Briefly, fNIRS signals were recorded using NIRx system (NIRScout, NIRx Medical Technologies, LLC) at wavelengths of 760 nm and 830 nm. The signals were sampled 10.41Hz. 16 NIR light sources and 24 detectors, forming 50 channels, were placed over the prefrontal and the motor cortices. Two mental arithmetic tasks (mental subtraction and backward counting) were considered. Each trial of the experiment consisted of 6 seconds of task interval, followed by a rest interval of $10-12$ seconds. Each block consisted of 13 trials of each task. During the experiment, 5 pain-free and 5 under-pain blocks were presented in random order, resulting in a total of 65 trials per task for each of the pain-free and under-pain conditions [16].

For the blocks that were accompanied with pain, subjects were exposed to thermal pain that was induced by applying heat to the dorsum of the left hand via a standard 30×30 thermode from TSA-II from Medoc. For each subject, the pain threshold was measured before applying the thermode. The process is described in [16]. For the pain-free blocks, the temperature of the thermode was set to baseline.

II-B. *Data preprocessing*

Pre-processing of the recorded signals was performed using nirsLAB. Signals were bandpass filtered between 0.01 to 0.2 Hz to remove artifacts, cardiac signal and low-frequency oscillations, and then converted into oxygenated ($[\Delta HbO_2]$) and deoxygenated ($[\Delta HbR]$) hemoglobin concentration changes using the modified Beer-Lambert's law [17]. The duration of $[-1$ to $6]$ seconds, with 0 indicating the onset of stimulus, was selected for each trial. $[\Delta HbO_2]$ signals were used due to their better signal-to-noise ratio compared to $[\Delta HbR]$ signals.

II-C. *Classification of tasks using CNN*

Recently, CNN has been explored for classification of fNIRS signals in various studies [18]-[23]. The architecture of the CNN used for classification of the tasks in the presence and the absence of pain is shown in Figure 1. The CNN was composed of 3 convolutional layers with 128, 64 and 32 filters and kernel sizes of 16, 8 and 4, respectively. Each convolution layer was followed by a dropout of 0.7 to minimize overfitting. The output from the third convolution layer after dropout, was fed to a maxpooling layer to reduce the size of features by downsampling the generated feature map by a factor of 4. The output from this layer was flattened and provided as input to a dense layer with 1 output for classifying the tasks. Rectified linear unit (ReLU) was used as the activation function for the convolution layers due to its least susceptibility to vanishing gradients. Sigmoid activation function was used for the output layer as it is non-linear, differentiable, and has an output range from 0 to 1, making it ideal for binary classification. The weight of the nodes in the network were updated using the adam optimizer. A learning rate of 0.0001 and exponential decay rate of 0.9 and 0.999 for the first and second moment estimates were used, respectively. The network was trained for a total of 500 epochs with a batch size of 32. Binary cross-entropy was used as the loss function for the optimizer. The CNN architecture in this work was chosen empirically. Future works will explore architecture selection methods for classification.

Besides adding dropout layers in the network, data augmentation, which is a method used to prevent overfitting in small datasets [24], was also employed. Specifically, the preprocessed signals were first resampled at 200 Hz, and signals from the interval of $[0-6]$ seconds were divided into 0.5-second segments, and were fed as inputs to the CNN with a total of 780 training samples (65 trials $\times$ 12 segments). Thus, the input data size was 780 (samples) $\times$ 200 (datapoints) $\times$ 50 (channels) per task. To account for the stochastic nature of the CNN, the training and testing processes were repeated 10 times. The average over 10 iterations was reported as the result.

We considered three cases to train and test the CNN classifier for the BCI.

Case 1 (NP-NP): The pain free data (no-pain (NP) data) was used for both training and testing the classifier. This situation occurs when the patient does not experience any pain during training or while using the BCI for assistive purposes.

Case 2 (WP-WP): The data obtained in the presence of pain (with-pain (WP) data) was used for both training and testing phases. This is the situation when the patient experiences pain of acute nature during training as well as when using the BCI.

Table 1. Labeling Schema for Multi-label Classification

Data	Pain Label	Task Label
Subtraction - No pain	0	0
Counting - No pain	0	1
Subtraction - With pain	1	0
Counting - With Pain	1	1

Table 2. Classification Accuracy (%) of the CNN

Subject	NP-NP	WP-WP	NP-WP
Subject 1	92.76±2.13	90.45±1.60	52.14±2.68
Subject 2	92.97±1.52	93.59±2.04	49.42±1.56
Subject 3	95.17±0.81	93.61±1.30	59.19±2.18
AVERAGE	93.63±1.48	92.55±1.65	53.58±2.15

Case 3 (NP-WP): The pain-free data (NP) was used for the training phase and the classifier was tested on data obtained in the presence of pain (WP). This case resembles the situation where the patient does not experience any pain while training for the BCI, but experiences acute pain while using the BCI.

The classification accuracy of the CNN is taken as the metric of performance for each case.

II-D. Multi-label classification

To further analyze the impact of the presence of pain for task classification, multi-label classification was performed on the data. In multi-label classification, a data sample can belong to more than one label at a time [25]-[28].

We assigned two labels to the fNIRS data. The first label indicated whether the data was obtained in the presence or the absence of pain, and the second label classified the task as either mental arithmetic or backward counting (see Table 1). For the pain label, the presence of pain was indicated by 1, and the absence of pain was indicated by 0. For the task label, the subtraction and the counting tasks were assigned 0 and 1, respectively. Multi-label classification was performed on the data using the same CNN architecture described in II-C. However, the number of outputs from the final dense layer was modified to 2 (one for each label). Since each label has a binary classification target, the sigmoid function was chosen for output activation.

We used the Hamming loss and micro-averaged F1 score as metrics for multi-label classification [28], [29]. The Hamming loss measures the proportion of incorrectly predicted labels to the total number of labels, and is expressed as

$$Hamming\ Loss = \frac{1}{nL}\sum_{i=1}^{n}\sum_{j=1}^{L} y\_pred_{i,j} == y\_gt_{i,j}, \quad (1)$$

where n is the number of data samples, L is the number of labels, *y_pred* is the predicted value, and *y_gt* is the ground truth of the value. The comparison is assigned a 0 if the predicted and the ground truth labels match, and 1 is assigned when there is a mismatch. Lower values of the Hamming loss indicate a better performance. The micro-averaged F1 score is the harmonic mean of the precision and recall scores of the data [28], is described as

$$Micro-averaged\ F1\ score = \frac{\sum_{i=0}^{1}(2*TP)_i}{\sum_{i=0}^{1}((2*TP)_i + (FP)_i + (FN)_i)} \quad (2)$$

where i represents the binary classification for each label, and *TP*, *TN*, *FP* and *FN* denote the true positives, true negatives, false positives and the false negatives, respectively. The micro-averaged score essentially computes the proportion of correctly classified observations out of all the observations. Thus, higher values of F1 scores indicate a better performance of the classifier.

II-E. Pooled data analysis

To mitigate the impact of pain on the classification accuracy of the BCI, we propose to train the classifier on both the data obtained in the presence and the absence of pain. This enables the BCI to learn the cortical signatures of the pain along with that of the tasks during the training phase. The cortical signals of the two tasks, collected both in the presence and the absence of pain, were pooled together and fed as input to the classifier during training. The trained classifier is then tested using NP data, and using WP data. The CNN model architecture described in section II-C is used for pooled data analysis as well.

III. Results

The classification accuracy of the CNN for the mental arithmetic tasks, for the three cases of NP-NP, WP-WP and NP-WP as described in II-C, are presented in Table 2. For case 1 (NP-NP), the cortical activities predominantly represent the tasks during both training and testing phases. The CNN model is seen to perform well with an average accuracy of 93.63% across the subjects. For case 2 (WP-WP), the BCI has assimilated the signature of pain along with cortical activity of the task during training. Since the nature of the pain experienced by the patient while using the BCI is similar as that experienced during the training phase, not much difference is expected in the cortical signals between the training and testing phases. As a result, a high classification accuracy is expected. Accordingly, in this case, the CNN yields an accuracy of 92.55%. In the third case (NP-WP), as the patterns of cerebral activity induced by pain is unknown to the model during the testing phase, the classification accuracy is expected to be lower. Accordingly, the accuracy drops to almost the chance level of 53.58%. This suggests that a BCI model trained with no knowledge of the pain signatures will fail to perform if the patient experiences pain later while using the BCI, leading to the potential failure of the BCI-controlled assistive device.

Table 3. Multi-label Classification Metrics

Subject	NP Data			WP Data		
	Hamming Loss	Micro-F1-Pain	Micro-F1-Task	Hamming Loss	Micro-F1-Pain	Micro-F1-Task
Subject 1	0.15	0.77	0.93	0.24	0.77	0.76
Subject 2	0.048	0.95	0.95	0.19	0.82	0.81
Subject 3	0.048	0.95	0.96	0.15	0.83	0.87
AVERAGE	0.083	0.89	0.95	0.19	0.80	0.81

For multi-label classification, the Hamming loss and micro-averaged F1 score metrics are presented in Table 3. The F1 score metrics is reported for the pain classification label (Micro-F1-Pain) and task classification label (Micro-F1-Task), separately.

The Hamming loss measures how well the classifier can make correct prediction on both the presence of pain and the task identification, simultaneously. It is observed that the average Hamming loss across all the subjects for the data obtained in absence of pain is 0.083, suggesting that the classifier can indicate the absence of pain and identify the task correctly at the same time for roughly 91.7% of the times $(1-0.083=0.917)$. However, the Hamming loss for data obtained in the presence of pain is about one order of magnitude higher than that for no-pain data. This means that the classifier can identify the presence of pain and the tasks performed in the presence of pain, simultaneously for only about 81% of the times $(1-0.19=0.81)$.

The micro-averaged F1 score for pain identification, averaged across subjects, is 0.89 for no-pain data and 0.80 for data obtained in the presence of pain. This represents the accuracy with which no pain and pain conditions are identified, respectively, while performing the tasks. Thus, the accuracy for pain identification while performing a task is 89% for no-pain and 80% for the presence of pain. These scores leads us to hypothesize that the classifier finds it challenging to isolate the pain component in the cortical signals while a patient is performing a task. The micro-averaged F1 score for task identification for no-pain data is 0.95, which represents the ability of the classifier to identify the tasks in the absence of pain. However, the micro-averaged F1 score for task identification in the presence of pain is 0.81. This represents the ability of the classifier to distinguish the two tasks while the subject is experiencing pain. The lowered accuracy of 81% compared to the no-pain data accuracy of 95% confirms that the task and pain signals from the brain cannot be isolated easily. It also reinforces that pain has a significant impact on the cortical signals of the task.

The pooled-data training method is used to mitigate the negative effect of pain on the performance of the BCI. The results are shown in Table 4. When testing the model on NP data, it yields an average accuracy of 94.64% across all the subjects. When the classifier is tested on WP data, the classification accuracy of the CNN is 83.67%. It can be seen that compared to the results shown in Table 2, this method of training has increased the accuracy for the NP-WP situation from 53.58%. It is interesting to see that the accuracy in the case of testing on WP data is lower than that of NP data. This can be attributed to the significant impact of pain on the cortical signals of the task and the inability of the classifier to distinguish the task component from the pain component in the cortical signals. However, this method of mitigation provides classification accuracy results that are high enough to be applied on practical binary classification BCIs that require a threshold classification accuracy of at least 70% [30].

Table 4. Classification Accuracy (%) of Pooled-data Training Method

Subject	NP	WP
Subject 1	91.86±2.44	79.53±3.59
Subject 2	95.88±1.48	82.92±2.42
Subject 3	96.20±1.28	88.55±2.0
AVERAGE	94.64±1.73	83.67±2.67

IV. Conclusions

In this work, we first investigated the impact of the presence of acute pain on the classification accuracy of mental arithmetic tasks in fNIRS-based BCIs, when CNN is used as the classifier. Our results indicated that while CNN provided a reliable way for classification of the tasks with greater than 90% accuracy, it performed poorly when the model is trained on pain-free data but tested on data collected in the presence of pain. To delve deeper into the analysis, we performed multi-label classification to detect the presence of pain and classify the task, simultaneously. The Hamming loss and micro-averaged F1 score metrics for this classification were better for no-pain data. The classifier performance was lower when it tried to detect the presence of pain and task, simultaneously. This leads to the conclusion that the cortical signals of the task and pain cannot be isolated easily and thus, pain has a significant impact on the cortical activity of the task at hand. Finally, to remedy the impact of pain, it was shown that the model can be trained collectively on task data obtained both in the presence and the absence of pain. In conclusion, it is of great importance to consider the presence of pain prior to adapting BCIs for assistive systems. Future work will explore methods to identify models that are not impacted by the presence of pain.

Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant No.1841087. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

References

[1] F. Shamsi, A. Haddad, and L. Najafizadeh, "Early classification of motor tasks using dynamic functional connectivity graphs from EEG," *Journal of Neural Engineering*, vol. 18, no. 1, 2021.

[2] M. Peifer, L. Zhu, and L. Najafizadeh, "Real-time classification of actual vs imagery finger tapping using fNIRS," *Biomedical Optics*. Optical Society of America, 2014, pp. 3A–34.

[3] F. Shamsi and L. Najafizadeh, "Multi-class classification of motor execution tasks using fNIRS," *IEEE Signal Processing in Medicine and Biology Symposium (SPMB)*, 2019, pp. 1–5.

[4] L. Zhu, S. Haghani, and L. Najafizadeh, "On fractality of functional near-infrared spectroscopy signals: analysis and applications," *Neurophotonics*, vol. 7, no. 2, p. 025001, 2020.

[5] F. Shamsi and L. Najafizadeh, "Multi-class fNIRS Classification of Motor Execution Tasks with Application to Brain-Computer Interfaces," *Biomedical Signal Processing*. Springer, 2021, pp. 1–32.

[6] J. E. Brown, N. Chatterjee, J. Younger, and S. Mackey, "Towards a physiology-based measure of pain: patterns of human brain activity distinguish painful from non-painful thermal stimulation," *PLoS ONE*, vol. 6, no. 9, p. e24124, 2011.

[7] R. F. Rojas, X. Huang, J. Romero, and K.-L. Ou, "fNIRS approach to pain assessment for non-verbal patients," *International Conference on Neural Information Processing*, 2017, pp. 778–787.

[8] R. F. Rojas, X. Huang, and K.-L. Ou, "A machine learning approach for the identification of a biomarker of human pain using fNIRS," *Scientific Reports*, vol. 9, no. 1, pp. 1–12, 2019.

[9] D. Lopez-Martinez, K. Peng, A. Lee, D. Borsook, and R. Picard, "Pain detection with fNIRS-measured brain signals: a personalized machine learning approach using the wavelet transform and Bayesian hierarchical modeling with Dirichlet process priors," *8th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW)*, 2019, pp. 304–309.

[10] R. F. Rojas, J. Romero, J. Lopez-Aparicio, and K.-L. Ou, "Pain assessment based on fNIRS using Bi-LSTM RNNs," *2021 10th International IEEE/EMBS Conference on Neural Engineering (NER)*, 2021, pp. 399–402.

[11] P. Rainville, "Brain mechanisms of pain affect and pain modulation," *Current opinion in neurobiology*, vol. 12, no. 2, pp. 195–204, 2002.

[12] D. A. Seminowicz and M. Moayedi, "The dorsolateral prefrontal cortex in acute and chronic pain," *The Journal of Pain*, vol. 18, no. 9, pp. 1027–1035, 2017.

[13] F. Shamsi and L. Najafizadeh, "On the effects of pain on fNIRS classification," *Optics and the Brain*. Optical Society of America, 2020, pp. 4C–6.

[14] F. Shamsi, A. Haddad, and L. Najafizadeh, "Recognizing pain in motor imagery EEG recordings using dynamic functional connectivity graphs," *42nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2020, pp. 2869–2872.

[15] T. Zeng, D. Peru, V. P. Maloney, and L. Najafizadeh, "Cortical activity changes as related to oral irritation-an fNIRS study," *39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2017, pp. 2558–2561.

[16] A. Subramanian, F. Shamsi, and L. Najafizadeh, "On the reliability of frequency-domain features for fNIRS BCIs in the presence of pain," *2021 IEEE Signal Processing in Medicine and Biology Symposium (SPMB)*, 2021, pp. 1–5.

[17] Y. Ardeshirpour, A. H. Gandjbakhche, and L. Najafizadeh, "Biophotonics techniques for structural and functional imaging, in vivo," *Stud Health Technol Inform*, vol. 185, pp. 265–97, 2013.

[18] U. Asgher, K. Khalil, Y. Ayaz, R. Ahmad, and M. J. Khan, "Classification of mental workload (MWL) using support vector machines (SVM) and convolutional neural networks (CNN)," *2020 3rd International Conference on Computing, Mathematics and Engineering Technologies (iCoMET)*, 2020, pp. 1–6.

[19] T. Ma, S. Wang, Y. Xia, X. Zhu, J. Evans, Y. Sun, and S. He, "CNN-based classification of fNIRS signals in motor imagery BCI system," *Journal of Neural Engineering*, vol. 18, no. 5, p. 056019, 2021.

[20] M. A. Tanveer, M. J. Khan, M. J. Qureshi, N. Naseer, and K.-S. Hong, "Enhanced drowsiness detection using deep learning: an fNIRS study," *IEEE access*, vol. 7, pp. 137 920–137 929, 2019.

[21] T. K. K. Ho, J. Gwak, C. M. Park, A. Khare, and J.-I. Song, "Deep leaning-based approach for mental workload discrimination from multi-channel fNIRS," *Recent trends in communication, computing, and electronics*. Springer, 2019, pp. 431–440.

[22] M. Saadati, J. Nelson, and H. Ayaz, "Mental workload classification from spatial representation of fnirs recordings using convolutional neural networks," *2019 IEEE 29th International Workshop on Machine Learning for Signal Processing (MLSP)*, 2019, pp. 1–6.

[23] H. Chhabra, N. Shajil, G. Venkatasubramanian *et al.*, "Investigation of deep convolutional neural network for classification of motor imagery fNIRS signals for BCI applications," *Biomedical Signal Processing and Control*, vol. 62, p. 102133, 2020.

[24] Z. Cui, W. Chen, and Y. Chen, "Multi-scale convolutional neural networks for time series classification," *arXiv preprint arXiv:1603.06995*, 2016.

[25] G. Tsoumakas and I. Katakis, "Multi-label classification: An overview," *International Journal of Data Warehousing and Mining (IJDWM)*, vol. 3, no. 3, pp. 1–13, 2007.

[26] K. Trohidis, G. Tsoumakas, G. Kalliris, I. P. Vlahavas *et al.*, "Multi-label classification of music into emotions." *ISMIR*, vol. 8, 2008, pp. 325–330.

[27] K. Sozykin, S. Protasov, A. Khan, R. Hussain, and J. Lee, "Multi-label class-imbalanced action recognition in hockey videos via 3d convolutional neural networks," *2018 19th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD)*. IEEE, 2018, pp. 146–151.

[28] N. M. Sommer, B. Kakillioglu, T. Grant, S. Velipasalar, and L. Hirshfield, "Classification of fnirs finger tapping data with multi-labeling and deep learning," *IEEE Sensors Journal*, vol. 21, no. 21, pp. 24 558–24 569, 2021.

[29] P. Kalansooriya, G. D. Ganepola, and T. Thalagala, "Affective gaming in real-time emotion detection and smart computing music emotion recognition: Implementation approach with electroencephalogram," *2020 International Research Conference on Smart Computing and Systems Engineering (SCSE)*. IEEE, 2020, pp. 111–116.

[30] C. Vidaurre and B. Blankertz, "Towards a cure for BCI illiteracy," *Brain topography*, vol. 23, no. 2, pp. 194–198, 2010.