Proceedings of the Fifth International Brain-Computer Interface Meeting 2013 DOI:10.3217/978-3-85125-260-6-111

Self-Paced BCI With NIRS Based on Speech Activity

C. Herff[1], D. Heger[1], F. Putze[1], C. Guan[2], T. Schultz[1]

[1]CSL, Karlsruhe Institute of Technology, Germany; [2]Institute for Infocomm Research, Singapore

Correspondence: C. Herff, Karlsruhe Institute of Technology, Institute for Anthropomatics, Cognitive Systems Lab (CSL), Adenauerring 4 (50.21), 76131 Karlsruhe, Germany. E-mail: christian.herff@kit.edu

Abstract. To enable Brain Computer Interfaces (BCIs) to be used intuitively, they should use an input paradigm which is as natural as possible, while the usage of the device is as convenient as possible. In this study, we show that functional near infrared spectroscopy (fNIRS) signals can be used to automatically detect the user's intent to use the system by detecting asynchronous speech activity, which is a very natural form of communication. Thereby, we take a first step towards self-paced BCIs with fNIRS based on speech activity.

Keywords: fNIRS, near infrared spectroscopy, asynchronous, self-paced, speaking modes

1. Introduction

Functional near infrared spectroscopy (fNIRS) is rapidly gaining attention as an imaging modality for Brain Computer Interfaces (BCIs) [Matthews et al., 2008]. First commercial systems are already available [Naito et al., 2007]. The majority of studies on fNIRS based BCIs rely on motor imagery and are operated stimulus locked. This means that the user can only interact with the system in predefined intervals. For an intuitive BCI, the users should be able to decide on their own when they want to operate the BCI. This is leading to a self-paced BCI, detecting idle and voluntary control states. Identifying segments containing activity is a known field of research in speech technologies [Laskowski and Schultz, 2006]. In this study, we show that the fNIRS signals measured during different speaking tasks can be automatically segmented into segments containing speech activity and segments without speech activity.

2. Material and Methods

We recorded 5 male subjects using a Dynot232 system with 32 transmitters and 32 receivers, sampling at 1.81 Hz. On the left hemisphere, four optodes were placed on Broca's area, 10 on Wernicke's area and six on 6 on the lower motor cortex. Additionally, we covered the prefrontal cortex with 12 optodes. Limiting our analysis to channels with an inter-optode distance between 2.5 and 4.5 cm, we considered 252 channels of oxygenated and deoxygenated hemoglobin values. All subjects were recorded over an interval of 37.5 minutes in which they conducted three types of speech activity as prompted by messages displayed on a screen. These were: Normal audible speech; silent speech, which consisted of moving the articulatory muscles as if speaking, but without sound production and speech imagery, for which the subjects had to imagine themselves reading out the displayed sentence. Users were asked to relax when they were not prompted to conduct speech activity. See [Herff et al., 2012b] for more details on the experiment design. The continuous hemoglobin values were then dissected into almost completely overlapping 10 second long windows, allowing for continuous decoding. A simple feature, measuring the difference between the mean of the first 4.5 s and the second 4.5 s in every window, was extracted for oxygenated and deoxygenated hemoglobin of every channel resulting in a total of 504 features. Using the *Mutual Information based Best Individual Feature (MIBIF)* algorithm [Ang et al., 2008], we selected a subset of 50 features which contained the most relevant information.

The data was labeled as containing speech activity (of any of the three modes) or not containing speech activity based on the experiment timings. This approach yielded more reliable results than identifying the modes individually, as more training data for the classes is available. Nevertheless, previous results [Herff et al., 2012b] show that all three modes can be discriminated from inactivity. We trained Support Vector Machines on these selected features to create an automated segmentation method for speech activity based on the fNIRS data.

A 10 fold cross-validation approach was used to evaluate our segmentation method. Both training and test set features were z-normalized with the mean and standard deviation of the training set. No data from the test set was used for feature selection, normalization or training of the classifier.

3. Results

Fig. 1 shows an example segmentation of our method for one of the folds of subject 3. The cross-marked line shows speech activity as labeled using the experiment prompts and the red line shows speech segments predicted by the proposed algorithm. Note that the ground truth might be flawed as well, since we could not control whether our subjects were actually performing the speech tasks when prompted. This typical example clearly shows that our

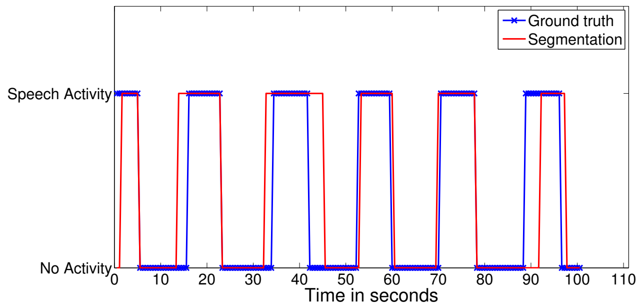

Figure 1: *Segmentation of subject 3's data into speech activity and non-activity. The cross-marked line is ground truth while the red line shows segmentation by our method.*

method extracts most of the segments containing speech activity reliably. The performance is promising on all 5 subjects and is significantly better than chance ($p < 0.05$). Precision and recall are very stable over all subjects and false positive rates are low across all subjects, as well. Frame based accuracies are high with an average of 74 % and are only slightly lower than the 79 % average accuracy achieved in a stimulus locked experiment on the same dataset [Herff et al., 2012b]. Table 1 lists all results for all subjects.

Table 1: *Results for speech activity detection with fNIRS across 5 subjects.*

Subject	Accuracy	True Positive Rate	False Positive Rate	True Negative Rate	Precision	Recall
Subject 1	0.72	0.61	0.20	0.80	0.69	0.61
Subject 2	0.74	0.62	0.17	0.83	0.71	0.62
Subject 3	0.79	0.63	0.11	0.89	0.79	0.63
Subject 4	0.73	0.57	0.16	0.84	0.72	0.57
Subject 5	0.74	0.63	0.17	0.83	0.72	0.63
Average	**0.74**	**0.61**	**0.16**	**0.84**	**0.73**	**0.61**

4. Discussion

We have shown that segments containing speech activity can be continuously decoded from those not containing speech activity based on fNIRS signals alone. Thereby, we make an important step towards self-paced BCI with fNIRS. All methods used in this analysis can be easily transferred to an online scenario. The results in this first study can possibly be extended to make use of the cross-subject capabilities of fNIRS [Herff et al., 2012a].

Acknowledgments

This project received financial support by the "Concept for the Future" of Karlsruhe Institute of Technology within the framework of the German Excellence Initiative.

References

Ang, K. K., Chin, Z. Y., Zhang, H., and Guan, C. (2008). Filter bank common spatial pattern (FBCSP) in brain-computer interface. In *IEEE International Joint Conference on Neural Networks. IJCNN*, pages 2390–2397. IEEE.

Herff, C., Heger, D., Putze, F., Guan, C., and Schultz, T. (2012a). Cross-subject classification of speaking modes using fnirs. In Huang, T., Zeng, Z., Li, C., and Leung, C., editors, *Neural Information Processing*, volume 7664 of *Lecture Notes in Computer Science*, pages 417–424. Springer Berlin Heidelberg.

Herff, C., Putze, F., Heger, D., Guan, C., and Schultz, T. (2012b). Speaking mode recognition from functional near infrared spectroscopy. In *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE*, pages 1715 –1718.

Laskowski, K. and Schultz, T. (2006). Unsupervised learning of overlapped speech model parameters for multichannel speech activity detection in meetings. In *Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on*, volume 1, pages I–I. IEEE.

Matthews, F., Pearlmutter, B., and Ward, T. (2008). Hemodynamics for brain-computer interfaces. *Signal Processing*, pages 87–94.

Naito, M., Michioka, Y., Ozawa, K., Ito, Y., Kiguchi, M., and Kanazawa, T. (2007). A communication means for totally locked-in als patients based on changes in cerebral blood volume measured with near-infrared light. *IEICE - Trans. Inf. Syst.*, pages 1028–1037.