Cross-Subject Classification of Speaking Modes Using fNIRS

Christian Herff[1,⋆], Dominic Heger[1], Felix Putze[1], Cuntai Guan[2], and Tanja Schultz[1]

[1] Cognitive Systems Lab (CSL),
Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany
{christian.herff,dominic.heger,felix.putze,tanja.schultz}@kit.edu
[2] Institute for Infocomm Research,
Agency for Science, Technology and Research (A*STAR), Singapore

Abstract. In Brain-Computer Interface (BCI) research, subject and session specific training data is usually used to ensure satisfying classification results. In this paper, we show that neural responses to different speaking tasks recorded with functional Near Infrared spectroscopy (fNIRS) are consistent enough across speakers to robustly classify speaking modes with models trained exclusively on other subjects. Our study thereby suggests that future fNIRS-based BCIs can be designed without time-consuming training, which, besides being cumbersome, might be impossible for users with disabilities. Accuracies of 71% and 61% were achieved in distinguishing segments containing overt speech and silent speech from segments in which subjects were not speaking, without using any of the subject's data for training. To rule out artifact contamination, we filtered the data rigorously.

To the best of our knowledge, there are no previous studies showing the zero training capability of fNIRS based BCIs.

Keywords: fNIRS, BCI, speech imagery, cross-subject, session-transfer.

1 Introduction

1.1 Motivation

A Brain-Computer Interface (BCI) is a communication channel between a user and a machine. Typical BCI applications target users with disabilities for whom standard input mechanisms are not feasible, due to motor limitations caused by brain stem stroke, cancer or amyotrophic lateral sclerosis, to name a few examples.

Functional Near Infrared Spectroscopy (fNIRS) provides robust measurement of hemodynamic responses in the brain, which are related to neural activity. It is

⋆ Part of this work was performed during the invited visit of the first author at A*STAR, Singapore, for which we are very thankful. This project received financial support by the 'Concept for the Future' of Karlsruhe Institute of Technology within the framework of the German Excellence Initiative.

T. Huang et al. (Eds.): ICONIP 2012, Part II, LNCS 7664, pp. 417–424, 2012.

less affected by artifacts caused by movements of the subjects than the de-facto standard modality in BCI, namely electroencephalography (EEG). Compared to functional magnetic resonance imaging (fMRI), which is based on the same hemodynamic effects, fNIRS is far cheaper and more portable. Even though fNIRS is a relatively new brain imaging modality, its feasibility for BCI has been shown in a number of papers [2,4].

Traditionally, BCIs rely on motor imagery for control, requiring the users to imagine movement of certain parts of their body. Naito et al. [11] first showed the usage of speech related activations, in the form of singing, with a very simple fNIRS sensor. In a very recent study [6], we showed that overt as well as imagined speech is a very reliable and promising paradigm for fNIRS-based human-machine interaction.

As brain signals are non-stationary and user-specific, i.e. they vary significantly over time and, even more so, between users, BCIs usually rely on training intervals from the same session to calibrate the system. Especially in applications for motion impaired users, a training procedure is cumbersome and reduces the time of actual interaction with the system. Recent advances in EEG-based BCI have shown that the usage of data from other subjects and sessions can reduce the time needed to calibrate the system [7,10] without compromising the system performance. With large numbers of sessions available for each subject, calibration time can completely be rendered obsolete [8].

In this study, we show that by using fNIRS data from other subjects, we can robustly distinguish between different speaking modes without any calibration data of the current user. Consequently, we do not require multiple sessions per user, but rely on only a very limited dataset of 5 subjects in total. Furthermore, our strict filtering assures that hemodynamic responses are used for classification while all artifacts are removed. The results achieved in this setup indicate the huge potential of fNIRS for BCIs, which are immediately usable without calibration time.

For this study, we investigated the following speaking modes in classification tasks: Normal audible speech (AUD_{Speech}), silently uttered speech, for which the subjects moved their articulatory muscles as if speaking but not producing any sounds (SIL_{Speech}), and speech imagery (IMG_{Speech}), for which the subjects had to to imagine themselves of speaking, including imagining to move their articulatory muscles.

1.2 Functional Near Infrared Spectroscopy

fNIRS measures the changes in oxy-hemoglobin (HbO) and deoxy-hemoglobin (HbR), which are triggered by changes in blood volume due to neural activity in the brain's cortical areas. Using light-sources and detector-optodes, which are fixated to the subjects' heads, these hemodynamic responses can be measured. Light in the near infrared range (620 - 1000 nm) disperses through biological tissue, such as scalp, skull and cortical areas of the brain, but is absorbed by hemoglobin. The modified Beer-Lambert law [12] can be applied to transfer raw optical densities (ΔOD) into changes in HbO and HbR, denoted as ΔHbO and

ΔHbR, respectively:

$$\Delta HbO = \frac{\Delta OD}{b \cdot l \cdot \alpha_{HbO}} \qquad \Delta HbR = \frac{\Delta OD}{b \cdot l \cdot \alpha_{HbR}} \tag{1}$$

with source-detector distance l, photon path length b and absorption coefficients α_{HbO} and α_{HbR} for HbO and HbR.

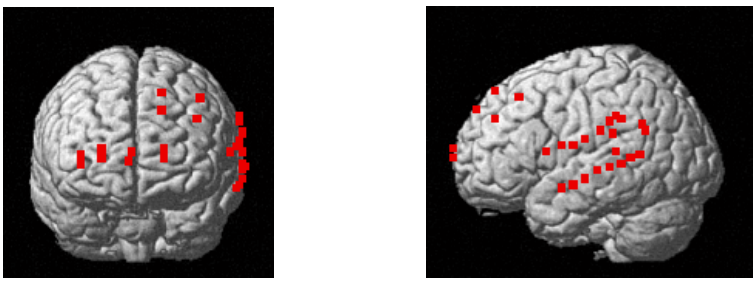
A typical hemodynamic response triggered by cortical activity increases on stimulus onset for HbO and decreases for HbR. After the end of the activation, the levels are expected to return to baseline.

2 Experiment

2.1 Setup

To record fNIRS data, we used a Dynot232 system by NIRX Medical Technologies equipped with 32 optodes, sampling at 1.81 Hz. All optodes were used as sources and detectors simultaneously. We used infrared wavelengths of 760 and 830 nm in this study. For every source-detector pair, the system outputs raw optical densities. We limited these to pairs with distances ranging from 2.5 to 4.5 cm, resulting in 252 channels of raw optical densities.

To measure neural activity in the relevant areas, four optodes were placed on Broca's area, 10 on Wernicke's area, both on the left hemisphere. The prefrontal cortex was covered with 12 optodes and six optodes were placed on the lower left motor cortex. Exact optode positions were registered with an ANT Visor infrared camera system[1] and plotted on a brain surface image using the NIRS-SPM software [13]. Figure 1 illustrates exact optode positions in our experiment.



(a) (b)

Fig. 1. (a) Optode positions frontal view. (b) Optode positions left lateral view. Created with [13].

[1] http://www.ant-neuro.com/products/visor/

2.2 Data Acquisition

Five male subjects participated in this study. All of them were right-handed and had a mean age of 27.6 years. Subjects had the 32 NIRS-optodes fixated to their heads by a helmet. Ten sentences in English from the broadcast-news domain were used for the experiment. Only subject 1's mother tongue was English, but all subjects spoke English fluently.

In the experiment, subjects produced utterances in the three modes AUD, SIL and IMG, where each utterance was separated by pauses. Sentences were prompted by displaying them on a screen placed 50cm away from the subjects. Trials are labeled according to the respective modes, i.e. AUD_{Speech}, SIL_{Speech}, IMG_{Speech} and AUD_{Pause}, SIL_{Pause}, IMG_{Pause}. In every mode, each sentence was repeated three times, resulting in a total of 30 trials per mode and per subject. Every utterance of a sentence and every subsequent pause are denoted as separate trials. Two subjects terminated the recordings prematurely resulting in fewer than 30 trials per mode. See Table 1 for full corpus characteristics.

The experimental design is described in more detail in our previous analysis [6].

Table 1. Corpus characteristics

Subject-ID	1	2	3	4	5
Mother tongue	English	German	Sinhala	German	Farsi
AUD_{Speech} trials	13	30	30	30	24
AUD_{Pause} trials	13	30	30	30	24
SIL_{Speech} trials	18	30	30	30	18
SIL_{Pause} trials	18	30	30	30	18
IMG_{Speech} trials	18	30	30	30	18
IMG_{Pause} trials	18	30	30	30	18
Total recording time (minutes)	20.6	37.5	37.5	37.5	25.2

3 Methods

3.1 Signal Preprocessing

The HomER package[2] was used to transfer the 252 channels of raw optical densities into ΔHbO and ΔHbR values. After linear detrending the channels, trials were extracted based on the experiment time information. Each trial was assigned a class label, which correspond to the *Speech* or *Pause* categories.

Cui et al. [5] showed that NIRS channels containing artifacts can be identified using the correlation between HbO and HbR. Usually, HbO and HbR should be strongly negatively correlated, but motion induced artifacts lead to positive correlations, as both values will spike when the optodes are shifted or are lifted off

[2] `http://www.nmr.mgh.harvard.edu/PMI/resources/homer/home.htm`

the scalp. To clean the data from artifacts, all channels which were not negatively correlated ($r > -0.3$) for every subject were removed from the dataset. This way, the initial 252 channels were reduced to 60 channels that do not contain artifacts for any of the subjects. Almost all channels on the forehead are removed through this procedure, as they are most vulnerable to movement induced artifacts.

3.2 Feature Extraction

Following Leamy et al. [9], we assume an idealized hemodynamic response for feature extraction. A rise in HbO is expected during speech activity and levels should return to baseline for the subsequent $Pause$ trials (and vice-versa for HbR). To make use of this observation, the mean μ of samples 9 to 15 (corresponding to roughly 4 seconds) is subtracted from the mean of the first 7 samples ($\sim$ 4 seconds) in every trial t for ΔHbO and ΔHbR for every channel i.

$$f_{i,t}^{\Delta HbO} = \mu(\Delta HbO_{t,1:7}^{i}) - \mu(\Delta HbO_{t,9:15}^{i}) \tag{2}$$

$$f_{i,t}^{\Delta HbR} = \mu(\Delta HbR_{t,1:7}^{i}) - \mu(\Delta HbR_{t,9:15}^{i}) \tag{3}$$

Given this feature extraction, we extract 120 features in total per trial. The features were normalized to zero mean and unit standard deviation (z-normalization).

3.3 Feature Selection

Ang et al. [1] presented the *Mutual Information based Best Individual Feature (MIBIF)* algorithm, a feature selection approach based on a high relevance criterion to reduce the feature space dimensionality. It has proven highly effective for BCI data [1] and is orders of magnitude faster than more complex *Mutual Information* based approaches which try to incorporate redundancy measures [3]. The Mutual Information $I(X;Y)$ can be understood as the amount of information shared by two random variables X and Y. Therefore, a feature containing highly relevant information should have a high Mutual Information with the class labels. *MIBIF* selects the k features with highest Mutual Information with the class labels. Assuming that the training data is representative of the test data, such selected features should increase the classification accuracy.

We set $k = 5$ after studying the distributions of Mutual Information of features with the class labels. See Figure 2 for the distribution of the Mutual Information when selecting features on four subjects for classification on the remaining fifth. Features are sorted decreasingly by their Mutual Information. It can be easily seen that the largest portion of the Mutual Information is explained by the first $k = 5$ features while the remaining 115 contribute only very little information. Selected features were very consistent across the different folds, but varied in between tasks.

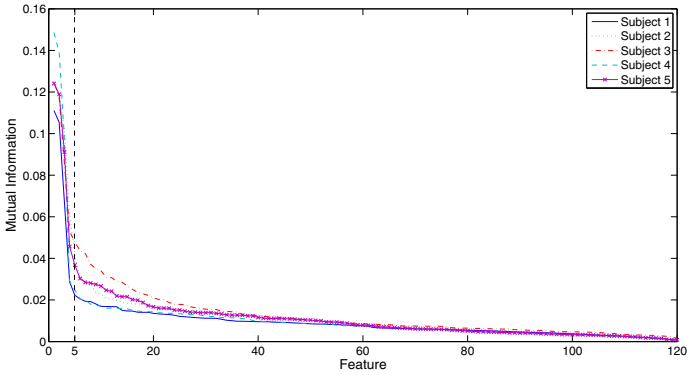


Fig. 2. Mutual Information over number of features for each subject when selecting features on the remaining four subjects for the AUD_{Speech} versus AUD_{Pause} task. The dotted line indicates the five selected features.

3.4 Classification and Evaluation

To evaluate our system, we applied a leave-one-speaker-out cross validation. A Linear Discriminant Analysis (LDA) classifier was trained on the 5-dimensional feature set S, determined with *MIBIF*. The LDA was trained on 4 subjects and tested on the remaining subject in a round-robin manner. Presented results were then averaged over all 5 rounds.

In a first experiment, all three *Speech* modes were combined and tested against all three combined *Pause* modes to discriminate speech activity from inactivity. Subsequently, every mode was classified from its respective *Pause* trials in binary classification experiments. Additionally, the three *Speech* modes were discriminated from each other.

4 Results

All classification results are presented in Figure 3. Differentiating between combined *Speech* (build from AUD_{Speech}, SIL_{Speech}, and IMG_{Speech}) and combined *Pause* worked reasonably well with an average accuracy of 58%. Subsequently, every *Speech* mode was tested individually against its respective *Pause* mode. Audible speech yielded best results with 71% average classification accuracy. This was expected, as neural activity from speech production, speech planning and auditory activity should be observed. Results for silent speech (SIL_{Speech}) are slightly lower (61%), which is explicable by the lack of auditory activity in the fNIRS signals. Discriminating IMG_{Speech} from IMG_{Pause}, when only speech planning activity is present, did not yield results better than chance level. This can be explained by the large variability in speech imagery across subjects, as their might be a lack of a consistent form of imagined speech, even though all speakers were instructed to imagine reading the sentences out loud.

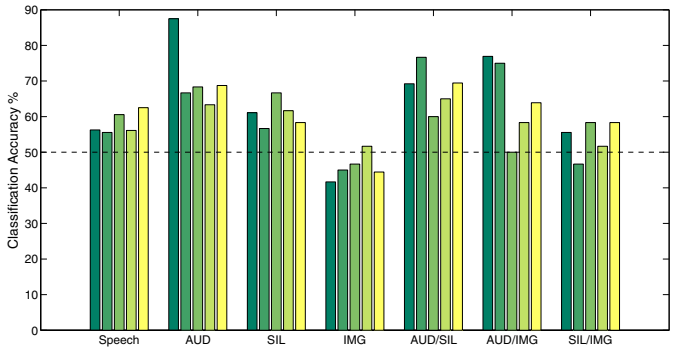


Fig. 3. Classification results for binary classification experiments *Speech* against *Pause* in all modes and between *Speech* of different speaking modes. Each color represents one subject. Dotted line stands for naive classification accuracies.

Our dataset is small and contains subjects from very different backgrounds (4 different mother tongues), thus the absence of a uniform activation pattern across subjects for speech imagery, for which neither muscle control, nor speech production or acoustic feedback are present is not too surprising.

Differentiating between the different speaking modes worked reliably, as well. Classification between AUD_{Speech} and SIL_{Speech} worked best with 68% accuracy. We were able to distinguish between AUD_{Speech} and IMG_{Speech} with 65% accuracy and our setup achieved 55% for SIL_{Speech} versus IMG_{Speech}.

In addition to the classification accuracies, we conducted t-tests to reject the null hypothesis that classification results were equal to naive classification. All experiments, except for IMG_{Speech} versus IMG_{Pause}, were significantly ($p < 0.05$) better than naive classification.

A summary of all classification results can be found in Table 2. These high results, which were achieved with the small dataset of just 5 subjects and which are rigorously artifact cleaned, show that fNIRS has huge potential for cross-subject classification in BCI.

Table 2. Average classification results and standard deviations in %

	Speech/*Pause*	AUD	SIL	IMG	AUD/SIL	AUD/IMG	SIL/IMG
Accuracy	58	71	61	46	68	65	54
Standard deviations	3.1	9.5	3.8	3.7	6.2	11.3	5.0

5 Conclusion

We have shown that fNIRS signals from speech related tasks produce brain activity that is consistent across multiple subjects. By selecting only the five

most relevant features that are reliable across all subjects, we are able to classify speaking modes solely based on training data from other subjects and thus make user specific training obsolete. Our rigorous filtering for artifacts and the significant results further support the argument that fNIRS signals from speech tasks have huge potential for future BCI applications, as they potentially reduce the amount of training needed in future experiments.

References

1. Ang, K.K., Chin, Z.Y., Zhang, H., Guan, C.: Filter bank common spatial pattern (FBCSP) in brain-computer interface. In: IEEE International Joint Conference on Neural Networks, IJCNN, pp. 2390–2397. IEEE (2008)
2. Ang, K.K., Guan, C., Lee, K., Lee, J.Q., Nioka, S., Chance, B.: A Brain-Computer Interface for mental arithmetic task from single-trial near-infrared spectroscopy brain signals. In: 20th International Conference on Pattern Recognition, pp. 3764–3767 (2010)
3. Battiti, R.: Using mutual information for selecting features in supervised neural net learning. IEEE Transactions on Neural Networks, 537–550 (1994)
4. Coyle, S.M., Ward, T.E., Markham, C.M.: Brain-computer interface using a simplified functional near-infrared spectroscopy system. Journal of Neural Engineering, 219–226 (2007)
5. Cui, X., Bray, S., Reiss, A.L.: Functional near infrared spectroscopy (NIRS) signal improvement based on negative correlation between oxygenated and deoxygenated hemoglobin dynamics. NeuroImage, 3039–3046 (2010)
6. Herff, C., Putze, F., Heger, D., Guan, C., Schultz, T.: Speaking mode recognition from functional near infrared spectroscopy. In: International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) (to appear, 2012)
7. Krauledat, M., Schröder, M., Blankertz, B., Müller, K.R.: Reducing calibration time for brain-computer interfaces: A clustering approach. In: Advances in Neural Information Processing Systems, pp. 753–760 (2007)
8. Krauledat, M., Tangermann, M., Blankertz, B.: Towards zero training for brain-computer interfacing. PLoS One, e2967 (2008)
9. Leamy, D.J., Collins, R., Ward, T.: Combining fNIRS and EEG to improve motor cortex activity classification during an imagined movement-based task. In: HCI (20), pp. 177–185 (2011)
10. Lotte, F., Guan, C.: Learning from other subjects helps reducing Brain-Computer Interface calibration time. In: IEEE International Conference on Acoustics Speech and Signal Processing, pp. 614–617 (2010)
11. Naito, M., Michioka, Y., Ozawa, K., Ito, Y., Kiguchi, M., Kanazawa, T.: 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., 1028–1037 (2007)
12. Sassaroli, A., Fantini, S.: Comment on the modified Beer-Lambert law for scattering media. Physics in Medicine and Biology, N255–N257 (2004)
13. Ye, J.C., Tak, S., Jang, K.E., Jung, J., Jang, J.: NIRS-SPM: Statistical parametric mapping for near-infrared spectroscopy. NeuroImage, 428–447 (2009)