Vocal Imagery vs Intention: Viability of Vocal Based EEG-BCI Paradigms

Kristensen, Alexander Borch; Subhi, Younes; Puthusserypady, Sadasivan

Published in:
IEEE Transactions on Neural Systems and Rehabilitation Engineering

Link to article, DOI:
10.1109/TNSRE.2020.3004924

Publication date:
2020

Document Version
Peer reviewed version

Citation (APA):
Vocal Imagery vs Intention: Viability of Vocal Based EEG-BCI Paradigms

Alexander Borch Kristensen, Younes Subhì, and Sadasivan Puthusserypady*, Senior Member, IEEE

Abstract—The viability of electroencephalogram (EEG) based vocal imagery (VIm) and vocal intention (VInt) Brain-Computer Interface (BCI) systems has been investigated in this study. Four different types of experimental tasks related to humming has been designed and exploited here. They are: (i) non-task specific (NTS), (ii) motor task (MT), (iii) VIm task, and (iv) VInt task. EEG signals from seventeen participants for each of these tasks were recorded from 16 electrode locations on the scalp and its features were extracted and analysed using common spatial pattern (CSP) filter. These features were subsequently fed into a support vector machine (SVM) classifier for classification. This analysis aimed to perform a binary classification, predicting whether the subject was performing one task or the other. Results from an extensive analysis showed a mean classification accuracy of 88.9% for VIm task and 91.1% for VInt task. This study clearly shows that VIm can be classified with ease and is a viable paradigm to integrate in BCIs. Such systems are not only useful for people with speech problems, but in general for people who use BCI systems to help them out in their everyday life, giving them another dimension of system control.

Index Terms—Brain Computer Interface, Electroencephalogram, Vocal Imagery, Vocal Intention, Support Vector Machine.

I. INTRODUCTION

Research within electroencephalogram (EEG) based brain computer interface (BCI) paradigms is advancing rapidly, providing new opportunities for EEG based applications [1]. BCI systems enable individuals for environmental control by mental thoughts, providing non-motor based interactions [2]. This is useful for people with neuropsychological and neurological disorders such as locked-in syndrome (LIS), stroke, amyotrophic lateral sclerosis (ALS), spinal cord injuries, traumatic brain injuries, etc., [3].

Imagery in cognition is a term used to refer to representations and experiences in the mind without external sensory stimulus. It is recalled from memory and is a mental recreation of an original stimulus or a novel combination of stimuli [4]. Various kinds of imagery have been studied and can generally be classified as either mental imagery or motor imagery (MI). Mental imagery refers to the experiences of sensory information and MI to the experiences of motor information. In MI, one mentally rehearses or simulates a given task and 91.1% for VInt task. This study clearly shows that VIm can be classified with ease and is a viable paradigm to integrate in BCIs. Such systems are not only useful for people with speech problems, but in general for people who use BCI systems to help them out in their everyday life, giving them another dimension of system control.

The first two authors contributed equally to this work. Asterisk indicates corresponding author.

B.K.Alexander is with the Department of Health Technology, Technical University of Denmark, Kgs. Lyngby 2800, Denmark (e-mail: s153675@student.dtu.dk).

YS.Subhì is with the Department of Health Technology, Technical University of Denmark, Kgs. Lyngby 2800, Denmark (e-mail: yssubhi@mit.edu).

S. Puthusserypady is with the Department of Health Technology, Technical University of Denmark, Kgs. Lyngby 2800, Denmark (e-mail: sapu@dtu.dk).

*EEG are associated with activation of neural circuits primarily in the pre-motor cortex, the primary motor cortex, and the supplementary motor area.

Imagery has been shown to be easy to study with a variety of modalities, such as functional near-infrared spectroscopy (fNIRS) [7], functional magnetic resonance imaging (fMRI) [8], magnetoencephalography (MEG) [9] and EEG [10]. The comparatively low complexity to study and classify imagery with techniques such as EEG, has opened for research in BCIs utilizing imagery as the sensory input (the brain signals) [11].

In the past decade, an increase in the development of EEG-based BCIs have been perceived due to the availability of powerful processing power at very low cost [1]. Several EEG-based BCI paradigms have been investigated intensively and MI based BCI has proven to provide high classification accuracy [1]. Steady state visual evoked potentials (SSVEP) is another popular paradigm with high accuracy, but lacks in convenience, requiring motor control of extraocular musculature, continuous visual attention, and external stimuli to produce the corresponding EEG signals [12]. The latter also applies to the P300 based BCI paradigm as well. However, one common short-coming of these (MI, SSVEP and P300) BCI paradigms is the limited range of commands that can be derived in a single system.

For the above-mentioned reasons and scientific curiosity, a novel humming (vocal imagery (VIm) and vocal intention (VInt)) based EEG-BCI paradigm is proposed in this work. Such a system would be useful for both people with speech problems as well as for people in general, who use BCI systems to help them out in their everyday life, giving them another dimension of system control.

Many researchers have exploited the use of speech imagery paradigms in EEG BCIs with varying success and one of the notable works has been by DaSilla et. al., in 2009. They have proposed a BCI control scheme using English vowel (/a/ and /u/) imagery to provide a speech prosthesis for severely communication impaired individuals [13]. In 2011, Chi et.al., have reported positive results on the classification of imagined phonemes using EEG signals and proposed the possibility of it in developing an online BCI application [14]. Leuthardt et al., in the same year, have demonstrated for the first time that the electrocortigogram (ECoG) signals associated with different overt and imagined phoneme articulation can enable invasively monitored human patients to control a 1D-computer cursor swiftly and precisely [15], Wang et. al., proposed combining speech imagery (based on Chinese characters) with MI in an EEG-based BCI for extending the possible number of control signals (operational dimensions) beyond 4 [16]. By combining three modalities (EEG, facial, and audio) during imagined and vocalized phonemic and single-word prompts, Zhao and Rudzicz have shown that it is possible to exploit multimodal relationships in developing BCI systems with high accuracies [17]. With the help of speech imagery signals, Wang et al., have showed that silent reading (of Chinese characters) can improve mental tasks for BCI systems [18]. In 2016, Min et. al., have proposed a vowel (/a/, /e/, /i/, /o/, and /u/) imagery decoding towards non-motor based interactions [2]. This is useful for people with speech problems, but in general for people who use BCI systems to help them out in their everyday life, giving them another dimension of system control.

The first two authors contributed equally to this work. Asterisk indicates corresponding author.

B.K.Alexander is with the Department of Health Technology, Technical University of Denmark, Kgs. Lyngby 2800, Denmark (e-mail: s153675@student.dtu.dk).

YS.Subhì is with the Department of Health Technology, Technical University of Denmark, Kgs. Lyngby 2800, Denmark (e-mail: yssubhi@mit.edu).

S. Puthusserypady is with the Department of Health Technology, Technical University of Denmark, Kgs. Lyngby 2800, Denmark (e-mail: sapu@dtu.dk).
with a single-trial EEG [19]. In the same year, Yoshimura et al., proposed EEG cortical currents as a novel approach for EEG-based BCI spellers for the communication impaired by decoding covert articulation of Japanese vowels /a/ and /i/ as well as during a no-imagery control task [20]. They have estimated the EEG cortical currents with a variational Bayesian method by using the fMRI data as a hierarchical prior. In another study, Brumberg et al., proposed a MI BCI system with real-time auditory and visual feedback of speech sounds (for production of three vowels, /i/, /a/, and /u/) that has the potential to serve as a communication tool for individuals with severe neuromotor impairments [21]. Most recently, Nguyen et al.,[22] combined MI and speech imagery for an online BCI control, to allow for tasks that involve multiple degrees of freedom. Dash et al., investigated speaker-independent neural speech decoding of five continuous phrases from MEG signals while the subjects produced speech covertly (imagination) or overtly (articulation) and showed significantly improved decoding performance in their method [23].

Same authors, in another study, have investigated automatic recognition of speech activity from MEG signals using machine/deep learning and reported accurate recognition of speech activity, which is critical for understanding the relationship between neural patterns and speech production [9] [24]. Though there are many studies on speech imagery based EEG BCI paradigms, Vlm is still considered a novel field without much published research works available. It differs from traditional MI, since it not only involves MI of speech articulating musculature, but also activates language and speech centers of the brain [25].

In this work, the classification accuracy of Vlm and Vln tasks were investigated from EEG, which could potentially be used as a viable paradigm in EEG-BCI systems. Seventeen participants performed three Vlm tasks based on humming. While, the first task involved actual vocal activity (motor action (MT)), the second task was pure Vlm, and the third task involved Vln. In a fourth set of recordings, the participants performed non-task specific (NTS) activity. We believe that the humming-based BCIs are interesting to investigate, due their potential as BCI systems with high levels of accuracy that require low levels of computational power. Further, the imagery tasks of humming have low thresholds for learning and making use of.

The remainder of this paper is divided into four sections structured as follows. In Section II, data and acquisition details as well as the different signal processing methods and their computations are described. Section III presents the experimental results, followed by a detailed discussion in Section IV. The concluding remarks are provided in Section V.

II. MATERIALS AND METHODS

A. Data acquisition

In our experiments, the EEG data was acquired and sent to MATLAB for processing, using a g.USBamp biosignal amplifier. These signals (from 16 electrode locations; scalp electrode locations will be described later) were recorded at 512Hz sampling rate and were bandpass (5-30Hz) filtered using a 8th order Butterworth filter. The ground electrode was placed on the forehead and the reference electrode on the left earlobe of the participant. The electrodes used were from the g.GAMMAbundle of the g.tec pack, used in conjunction with Nuprep skin prep gel, alcohol wipes and Sigma electrode gel.

1) Participant information: Seventeen subjects (11 males and 6 females and aged between 19 and 30 years with a mean age of 22.56 ± 1.73 years) participated in the study. All participants were healthy with no known neurological or psychiatric disorders. The participants provided written informed consent prior to their participation in the experiment, which were approved by the Regional Committee on Health Research Ethics for the Capital Region of Denmark (reference H-3-2013-004). The experiments were carried out in accordance with the corresponding guidelines and relevant regulations on the use of human participants for health-related scientific research. Data was stored according to the guidelines of The Danish Data Protection Agency and has been anonymized.

2) Recording protocol: EEG recordings were performed in a Faraday cage for noise reduction. Participants were seated on a comfortable office chair in front of a laptop and recording devices. Four different tasks were performed continuously by the participants during trials of one-minute long. The participants were instructed to initiate a trial recording by clicking start button when ready, and thereby able to control the length of breaks in between measurements. Each participant completed 10 trials of each task. Before starting a new task, the participants made a test-trial, which was discarded from the analysis. Post recording, these 10 trials were concatenated to yield one 10-minute long segment for each task per subject.

Prior to and after recordings, participants were instructed to answer a questionnaire, declaring their age, gender, concentration ability prior to recording (1-7), perceived concentration ability during recordings (1-7), concentration level after recordings (1-7) and supplementary comments.

3) Tasks: The four tasks in the order they performed were: NTS, sound production/motor action (MT), Vlm, and Vln. In the first task (NTS), participants were instructed not to think of anything specific and not to perform or imagine any form of vocal activation. In MT, participants were instructed to perform continuous humming. Natural breaks for breathing was allowed during the experiments. In the third task (Vlm), participants were instructed to focus on the tactile feeling of humming, without producing sound or activating related muscles. In the final and fourth task (Vln), participants were instructed to continuously develop intention of initiating vocal activation, without imagining the sound or the sound production. Eyes were kept open during each task, blinking was allowed as needed.

4) Electrode positions of EEG setup: Producing sound requires use of a series of muscular elements, the most important being the diaphragm, regulating the speed of exhalation, the vocal cord, tongue and oro-facial muscles. Vocal activity is produced through exhaling, while compressing the air in order to control the pitch. Further activation of certain centers involved in language and speech production is also anticipated. Therefore, it can be anticipated that the origin of the signals can be tracked back to infero-lateral most parts of the motor cortex equivalent to the infero-posterior frontal lobe.
and infero-anterior parietal lobe (Brodmann 43 and lower 1,2,3) for the motor signals and primarily pars opercularis of Broca’s area (Brodmann 44) for speech production [26]. The electrodes were placed (Fig. 1) at FC5, FC3, FC4, FC6, T7, C5, C3, C4, C6, T8, TP7, CP3, CP4, TP8, P3 & P4 positions on the scalp in accordance with the international 10-20 system [27]; covering the anticipated brain areas of interest.

B. Data analysis

Here, the different steps adopted to process the collected EEG signals is briefly explained.

1) Feature Extraction: The efficiency of a BCI system, at large, depends on the correct choice of features. An increase in the number of features could improve the efficiency; however, it also results in an increased risk of overfitting the classifier. While there is no one correct model of feature extraction in EEG-based recordings due to its subject-specific features, different methods of feature extraction have their respective advantages for specific recordings.

2) Common Spatial Pattern Filters: The Common Spatial Pattern (CSP) method, first employed by [28], was used for binary classification and spatial filtering.

The CSP method assumes that there is a weight matrix or projection matrix, \( \mathbf{W} \), such that the measured EEG channels, \( \mathbf{X} \), can be linearly transformed to ‘source channels’, \( \mathbf{S} \):

\[
\mathbf{S} = \mathbf{W}^T \mathbf{X},
\]

such that the variance of \( \mathbf{S} \) yield maximum separability between classes. Here, \( \mathbf{T} \) denotes the matrix transposition, and \( \mathbf{X} \) represents the collected EEG data where each row correspond to one channel. If \( \mathbf{W} \) exist, it can be determined through a multitude of methods.

In this investigation, the method of generalized eigen-decomposition (GED) has been applied for determining \( \mathbf{W} \).

The covariance matrix of signal \( \mathbf{X}_c \) for class \( c \), is defined as:

\[
\mathbf{C}_c = \frac{\mathbf{X}_c \mathbf{X}_c^T}{\text{trace}(\mathbf{X}_c \mathbf{X}_c^T)}.
\]

First the eigen-decomposition is performed on the joint covariance matrix of the 2 classes, determining \( \mathbf{U} \) (matrix of eigenvectors) and \( \mathbf{\Lambda}_a \) (matrix of eigenvalues):

\[
\mathbf{C}_1 + \mathbf{C}_2 = \mathbf{U} \mathbf{\Lambda}_a \mathbf{U}^T.
\]

Here, \( \mathbf{C}_1 \) and \( \mathbf{C}_2 \) are the covariance matrices for class 1 and 2, respectively. A whitening matrix, \( \mathbf{P} \), is then computed as:

\[
\mathbf{P} = \sqrt{\mathbf{\Lambda}_a^{-1}} \mathbf{U}^T.
\]

The weight matrix, \( \mathbf{W} \), can now be computed as:

\[
\mathbf{W} = \mathbf{P}^T \mathbf{V}.
\]

Here, columns of \( \mathbf{V} \) are eigenvectors for the covariance matrix of either class 1 or class 2, after the ‘whitening’ operation. For class 1, the eigen-decomposition would look like:

\[
\mathbf{P} \mathbf{C}_1 \mathbf{P}^T = \mathbf{V} \mathbf{\Lambda}_1 \mathbf{V}^T.
\]

When the weight matrix has been created, the EEG data can be transformed to source space according to Eq. (1). The variance of the sources are then used to compute the features, \( \mathbf{y} = [y_1, y_2, \ldots, y_k]^T \), where:

\[
y_i = \log \left( \frac{\text{var}[s_i]}{\sum_{j=1}^{k} \text{var}[s_j]} \right), \quad i = 1, 2, \ldots, k,
\]

where \( k \) refers to the number of sources (\( s_j \)'s) or rows in \( \mathbf{S} \) from Eq. (1). The rows of \( \mathbf{S} \) can be reduced by choosing only the first \( m(<k) \) last columns of the projection matrix in Eq. (1). This method for dimension reduction is commonly practised, but has not been done in the current work as the dimension reduction is not essential for the purpose of demonstrating Vlm and VInt tasks. The feature vector, \( \mathbf{y} \), is then fed into the support vector machine (SVM) for classification.

3) Support Vector Machine: The SVM is a classifier that finds the separating hyper-plane for which the margin between samples of the two classes are maximized. In a number of BCI applications, SVMs have been shown to be efficient in terms of classification accuracy, processing power, and its potential for easy transferability to an online BCI system [29]–[34]. For these reasons, we have chosen SVM in our study, and this method was implemented using the built-in \texttt{fitcsvm} and \texttt{predict} functions in MATLAB.

4) Epoch extraction: Feature vectors to train the classifier and for classification is made based on epochs extracted from the measured signal i.e. the steps in Eq. (1) and Eq. (7) is performed on the epochs. It is expected that making epochs too short will yield poor accuracy by the classifier. On the other hand, making the epochs too long will result in poor time resolution, if the system is to be implemented in a real-time BCI scenario. Suitable epoch length and overlap of neighbouring epochs is determined by trial-and-error.

C. Performance

With the future potential of using Vlm in BCI systems, accuracy is determined subject-wise as these systems can be personally tailored for the individual brain signal variability. Thus the model will also be able to adjust for variations in electrode impedance and sensitivity. The performance of the system was evaluated on binary classification to distinguish any combination of two tasks, e.g. NTS vs Vlm. This was measured in terms of the accuracy, which is defined as:

\[
\text{Accuracy} = \frac{TP + TN}{\text{Total Classifications}},
\]

where, \( TP \) and \( TN \) refers to the true positive and true negative classes, respectively. The accuracy was estimated using a 10-fold cross-validation method. The procedure of comparing two classes using the 10-fold cross-validation is described in a step-wise manner for clarity: itemsep=-0.5mm

1) The data is divided into 10 equally sized temporal segments.
2) For both classes to be compared, 9/10 segments are concatenated to yield one training set.
3) A weight matrix as described in Eqs. (2)-(6) is computed from the entire training set.
4) Epochs are extracted from the same training data and converted to source space described in subsection II-B.4. The resulting features are used for training a classifier.
5) Epochs are extracted from the last 1/10 segment of the data, to create feature vectors and to test the classifier.
6) Steps 1–5 is repeated using the next 1/10th of the data set, until all data has been used as test data.
7) The accuracy for the chosen subject is given as the average accuracy from the 10 tests.

Epochs overlapping the training and the test data was not regarded for that particular training/test.

D. Population average

Columns of the weight matrix, \( \mathbf{W} \), is used for creating topographic maps yielding high separability between the respective classes. The covariance as in Eq. (2) is found for each subject of the entire dataset.
The population covariance matrix for a class is then obtained by averaging across subjects as below:

$$C_c = \frac{1}{N_s} \sum_{i=1}^{N_s} C_{c,m},$$

(9)

where $C_c$ is the population covariance matrix of class $c$, and $N_s$ is the total number of subjects. Using the average covariance matrix instead when computing the weight matrix in Eqs. (3)-(6) will result in the topographic maps reflecting population filters.

**E. Spatial filters and spatial patterns**

An important distinction is to be made regarding spatial filters and patterns as this occasionally is the cause of confusion, when plotting topographic maps. Considering Eq. (1), the signal matrix, $X$, at the electrodes is a projection of source activity by the forward model:

$$A^T S = X,$$

(10)

where $A = W^{-1}$, is commonly referred to as the mixing- or the forward model [35]. Columns of $W$ is referred to as the spatial filters, while columns of $A$ are referred to as spatial patterns. While spatial filters yield the highest variance based separability between two classes, spatial patterns demonstrate how the estimated sources project onto the respective EEG-channels. Furthermore, the $i^{th}$ filter is in general designed to catch activity of the $i^{th}$ pattern in the columns of $A$ and $W$ are related by Kronecker’s delta function:

$$w_i^T a_j = \begin{cases} 
1, & \text{if } i = j, \\
0, & \text{otherwise}, 
\end{cases}$$

(11)

where $a_j$ is the $j^{th}$ column of $A$ and $w_i$ is the $i^{th}$ column of $W$ [36]. However, the appearance of the estimated spatial patterns depends on all columns of the weight matrix. With the expectation that some of the less important filters might vary by chance, this will effect all spatial patterns to some degree. We therefore choose to only represent the filters by topographic mapping for consistency.

**III. RESULTS**

In this section, a detailed analysis of the results obtained in our extensive studies are provided.

**A. Epoch size**

Accuracy dependency on epoch size is analyzed and demonstrated in Fig. 2 when classifying NTS vs VIm and NTS vs VInt. Based on this results, subsequent analyses are carried out on epochs of 600 samples, as only little improvement is seen beyond this epoch size, while maintaining a relative short time resolution. For 512 Hz sampling, this corresponds to a time resolution of 1.17 seconds.

**TABLE I:** Subject, binary classification accuracy from 10-fold cross-validation of the CSP procedure, with fixed epoch length of 600 samples.

<table>
<thead>
<tr>
<th>Subj#</th>
<th>NTS vs MT</th>
<th>NTS vs VIm</th>
<th>NTS vs VInt</th>
<th>MT vs VIm</th>
<th>MT vs VInt</th>
<th>VIm vs VInt</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Female Subjects</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>0.983</td>
<td>0.997</td>
<td>0.979</td>
</tr>
<tr>
<td>2</td>
<td>0.933</td>
<td>0.940</td>
<td>0.961</td>
<td>0.872</td>
<td>0.890</td>
<td>0.902</td>
</tr>
<tr>
<td>3</td>
<td>0.952</td>
<td>0.854</td>
<td>0.878</td>
<td>0.774</td>
<td>0.866</td>
<td>0.662</td>
</tr>
<tr>
<td>4</td>
<td>0.982</td>
<td>0.999</td>
<td>1.000</td>
<td>0.993</td>
<td>0.978</td>
<td>0.915</td>
</tr>
<tr>
<td>5</td>
<td>0.918</td>
<td>0.786</td>
<td>0.923</td>
<td>0.873</td>
<td>0.899</td>
<td>0.896</td>
</tr>
<tr>
<td>6</td>
<td>0.871</td>
<td>0.934</td>
<td>0.950</td>
<td>0.706</td>
<td>0.855</td>
<td>0.895</td>
</tr>
<tr>
<td>Avg.</td>
<td>0.942</td>
<td>0.919</td>
<td>0.952</td>
<td>0.868</td>
<td>0.914</td>
<td>0.844</td>
</tr>
<tr>
<td>std.</td>
<td>0.046</td>
<td>0.084</td>
<td>0.047</td>
<td>0.114</td>
<td>0.059</td>
<td>0.099</td>
</tr>
<tr>
<td></td>
<td>Male Subjects</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>0.789</td>
<td>0.774</td>
<td>0.838</td>
<td>0.745</td>
<td>0.782</td>
<td>0.646</td>
</tr>
<tr>
<td>8</td>
<td>0.948</td>
<td>0.941</td>
<td>0.968</td>
<td>0.801</td>
<td>0.829</td>
<td>0.656</td>
</tr>
<tr>
<td>9</td>
<td>0.894</td>
<td>0.910</td>
<td>0.936</td>
<td>0.878</td>
<td>0.903</td>
<td>0.564</td>
</tr>
<tr>
<td>10</td>
<td>0.926</td>
<td>0.983</td>
<td>0.992</td>
<td>0.918</td>
<td>0.955</td>
<td>0.850</td>
</tr>
<tr>
<td>11</td>
<td>0.951</td>
<td>0.933</td>
<td>0.934</td>
<td>0.927</td>
<td>0.890</td>
<td>0.891</td>
</tr>
<tr>
<td>12</td>
<td>0.926</td>
<td>0.941</td>
<td>0.958</td>
<td>0.955</td>
<td>0.961</td>
<td>0.743</td>
</tr>
<tr>
<td>13</td>
<td>0.989</td>
<td>0.633</td>
<td>0.798</td>
<td>0.956</td>
<td>0.928</td>
<td>0.671</td>
</tr>
<tr>
<td>14</td>
<td>0.946</td>
<td>0.863</td>
<td>0.942</td>
<td>0.948</td>
<td>0.963</td>
<td>0.732</td>
</tr>
<tr>
<td>15</td>
<td>0.864</td>
<td>0.702</td>
<td>0.638</td>
<td>0.676</td>
<td>0.727</td>
<td>0.380</td>
</tr>
<tr>
<td>16</td>
<td>0.905</td>
<td>0.982</td>
<td>0.966</td>
<td>0.847</td>
<td>0.847</td>
<td>0.757</td>
</tr>
<tr>
<td>17</td>
<td>0.959</td>
<td>0.943</td>
<td>0.907</td>
<td>0.963</td>
<td>0.928</td>
<td>0.786</td>
</tr>
<tr>
<td>Avg.</td>
<td>0.918</td>
<td>0.873</td>
<td>0.889</td>
<td>0.874</td>
<td>0.883</td>
<td>0.698</td>
</tr>
<tr>
<td>std.</td>
<td>0.055</td>
<td>0.118</td>
<td>0.107</td>
<td>0.097</td>
<td>0.078</td>
<td>0.141</td>
</tr>
<tr>
<td>Avg(all)</td>
<td>0.927</td>
<td>0.889</td>
<td>0.911</td>
<td>0.872</td>
<td>0.894</td>
<td>0.749</td>
</tr>
<tr>
<td>std</td>
<td>0.052</td>
<td>0.107</td>
<td>0.094</td>
<td>0.100</td>
<td>0.072</td>
<td>0.144</td>
</tr>
</tbody>
</table>

The equivalent Receiver Operator Characteristic (ROC) curves of the respective comparisons is illustrated in Fig. 3. Each point of an ROC is given as a mean of the equivalent point across the population. The equivalent Area Under the Curve (AUC) is used as a measure for the classification efficiency and is summarized in Table II. Each AUC is computed as the mean AUC as computed for each individual subject.

The AUC scores complies well with the accuracy scores for the whole population given in Table I. Demonstrating highest separability NTS vs MT, NTS vs VIm or MT vs VIm (AUC > 0.95). NTS vs VIm and MT vs VIm also show a high separability (0.94 > AUC > 0.92). Meanwhile, VIm has the worst separation from VInt (AUC = 0.812).

**C. Spatial filters and patterns**

Figure 4 demonstrates the first three ($1 − 3$) and last three ($14 − 16$) spatial filters of the respective data sets in comparison.
Each subject the PSD was found for all filters. By inspecting these clearer tendencies. Obscurities were discovered for individual sub-signal with the trace of its own covariance matrix, demonstrating on the PSD appearance.

Only three filters are included from each classification comparison, as the additional filters show no interesting results for discussion. Interestingly, many spatial filters are found to be common or similar, when comparing NTS to MT, VIm or VInt.

Note that the topographic plots extrapolate values for the entire scalp despite that distant areas contribute very little to the signal at the electrodes. The reader’s discretion is advised not to interpret too deeply from the distant cortical areas.

As mentioned, the CSP procedure finds spatial filters, that demonstrate the highest separability, based on variance. This means that one could take any single one of the filters multiply it by \(-1\), and the weight matrix still yields filters of same separability. In this case, the filter #16 in Fig. 4b is found to be the same but of opposite sign to filter #16 of Fig. 4a and 4c, by the procedure. It is the same for filter #3 in Fig. 4c compared to 4a and 4b. These cases are likely the result from activity of the same or similar sources.

1) Spectral analysis: It was discovered that by excluding single subjects from the population, filter estimation could shift the position in which the filters appear and some filters may completely change.

A filter related to NTS task is expected to demonstrate higher power for NTS rather than the opposing task(s). Thus creating Power Spectral Density (PSD) of the filtered signal should demonstrate clear difference in power between tasks if the filter reflects a legitimate source activity. PSDs of the filtered signal is used to find relevant filters and to illustrate band-limited source behavior. For better PSD estimates, each subject dataset is divided into 100 sub-segments, the PSD is averaged across these sub-segments and across subjects. The relevant filter-PSD plots for illustration were selected manually based on the PSD appearance.

All the PSDs were obtained (and shown) after normalizing the EEG signal with the trace of its own covariance matrix, demonstrating clearer tendencies. Obfuscations were discovered for individual subjects by first making filters using data from all subjects, then for each subject the PSD was found for all filters. By inspecting these individual spectra for each subject-filter combination, it was evident that some subjects has spectra of amplitudes that would dominate an average spectrum if they were included. Additionally, the spectra lacked systematic band activity which otherwise was evident from the rest of the subjects. From these criteria, subjects 4, 8, 12, and 16 were excluded from further PSD analysis.

As found in Fig. 4, some filters reappear comparing NTS against the other tasks. To better estimate these filters, an average covariance matrix is made for MT, VIm and VInt before finding \(\mathbf{W}\) in the CSP procedure. The two dominant filters are illustrated in Fig. 5a. Interestingly, the two maps seem to be almost identical but mirrored. The top spectrum demonstrates clear event related desynchronization (ERD) in the approximate frequency range [5–13]Hz, when performing MT, VIm or VInt. The bottom pattern demonstrates event related synchronization (ERS) of all the same tasks, primarily VIm and VInt, in a different band approximately [4–8]Hz. This band might even reach lower since these frequencies have been filtered out during data acquisition.

The top and middle filter of Fig. 5b also demonstrates a visible ERD band-limited to the frequency range [9.5–12.5]Hz for the MT, VIm and VInt tasks. The bottom filter in the same figure demonstrates high power of the MT, VIm and VInt however, this effect is spread over the entire frequency range of the applied analogue band-pass filter. Basis for this broad band behavior is unclear, but we believe it is caused by the filter catching the activity of multiple sources.

The accuracy scores where relatively high between MT and VInt tasks, indicating that the two tasks are driven by different sources.
of activities. To investigate these sources, the same procedure is conducted using population filters with high separability between the MT and VInt tasks. The resulting manually selected filter-PSD plots are illustrated in Fig. 6.

As can be seen in Fig. 6a bottom filter, #16, is the same as previously, when separating NTS from the other 3 tasks. This reflects a source that is related to MT but mostly active during a VIm or VInt state. Activity of the source is, potentially linked to restriction of motor actuation, thus demonstrating the largest power when motor activity is imagined or intended and lower activity for restricting the amplitude of the motor activity, during MT.

The top filter, #12, in Fig. 6a shows a band-limited ERD during MT and VInt task in the approximate band [5.5–12.5]Hz, primarily MT. Filter #3 at the bottom of Fig. 6b shows band-limited similar to the one seen previously in Fig. 5. Surprisingly, the VInt and the MT PSDs are almost identical. The top filter, #1, in Fig. 6b demonstrates a broadband increase in power during MT. Surprisingly, this is the opposite effect than VIm and VInt with respect to NTS.

D. Statistical analysis: t-test / P-value analysis

Statistical analysis is conducted to test if the performance of female subjects are different from that of male subjects. From the results presented in Table I, it appears that the female subjects in general perform better in terms of accuracy in any binary classification scenarios. T-tests are carried out to test for differences in accuracy scores between male and female subjects. P-values are reported testing the null-hypothesis ($H_0$) and the alternative hypothesis ($H_1$):

$$ H_0 : \text{The mean of accuracy scores for female and male subjects are the same, when classifying \{\text{task X}\} \text{ vs } \{\text{task Y}\}, and} $$

$$ H_1 : \text{The mean of accuracy scores is larger for female than male subjects, when classifying \{\text{task X}\} \text{ vs } \{\text{task Y}\}.} $$

For the $t$–test, the female and male accuracy scores are assumed to have equal but unknown variance. One of the criteria for conducting t-test is that the compared samples are from a normally distributed population. This criteria is known to be violated since accuracy scores are limited to a range, however approximate normality is acceptable. Prior to each $t$–test, a Kolmogorov-Smirnov (KS) test is performed to check for normality. If this cannot reject a null hypothesis that either the female or male accuracy is from a normally distributed with a significance level of $\alpha = 0.05$, the normality assumption is deemed acceptable for $t$–testing.

The p-values for the KS tests and the t-tests are summarized in Table III. From the table, it can be seen that none of the KS test leads to p-value below 0.05, meaning the normality assumption holds sufficiently.

The resulting p-values of the $t$–test cannot statistically infer that female subjects perform better in terms of accuracy than male subjects in any of the binary classification tasks.

E. Areas of brain activity

Increased brain activity in specific brain regions are expected in correlation with the different tasks. For all tasks, increased activity in frontal areas and language centers are expected. This is thought to be the result of conceptualizing and vocal activity. Likewise,
increased activity corresponding to the motor areas is expected as well. In order to localize the areas of brain activity, the topographic maps of CSP filters 1 – 3 and 14 – 16 were qualitatively analyzed. The topographic overview was split in to 10 sections corresponding to left/right frontal lobes, left/right motor cortex, left/right sensory cortex, language centers and left/right occipital lobes. In each map of CSP filters 1 – 3 and 14 – 16 for all participants, a review of activity was performed. High levels of activity for each area in all tasks across participants were counted and summed, showing recurrence of activity or lack thereof for each area. Results (refer Table IV) show high activity in Right Motor Cortex (RMC) and areas corresponding to Language Centers for MT task; Left Frontal Lobe (LFL), Left Language Centers (LLC), Right Frontal Lobe (RFL), RLC and Right Language Center (RLC) for VIm task; LFC, LLC, RFC and RMC for VInt task.

<table>
<thead>
<tr>
<th>Area of Activity</th>
<th>MT Task</th>
<th>VIm Task</th>
<th>VInt Task</th>
</tr>
</thead>
<tbody>
<tr>
<td>LFL High (72.5%)</td>
<td>High (70.6%)</td>
<td>High (76.5%)</td>
<td></td>
</tr>
<tr>
<td>LMC Average (65.7%)</td>
<td>Average (56.9%)</td>
<td>Average (51.8%)</td>
<td></td>
</tr>
<tr>
<td>LSC Low (29.7%)</td>
<td>Average (45.1%)</td>
<td>Average (53.9%)</td>
<td></td>
</tr>
<tr>
<td>LLC High (76.5%)</td>
<td>High (75.5%)</td>
<td>High (68.6%)</td>
<td></td>
</tr>
<tr>
<td>LOL Average (53.9%)</td>
<td>Average (41.2%)</td>
<td>Average (63.72%)</td>
<td></td>
</tr>
<tr>
<td>RFL High (77.5%)</td>
<td>High (67.6%)</td>
<td>High (82.6%)</td>
<td></td>
</tr>
<tr>
<td>RMC High (91.2%)</td>
<td>Average (64.7%)</td>
<td>High (70.6%)</td>
<td></td>
</tr>
<tr>
<td>RSC Low (24.2%)</td>
<td>Average (43.1%)</td>
<td>Average (54.9%)</td>
<td></td>
</tr>
<tr>
<td>RLC High (75.5%)</td>
<td>High (79.41%)</td>
<td>High (70.6%)</td>
<td></td>
</tr>
<tr>
<td>ROL Average (51.9%)</td>
<td>Average (50%)</td>
<td>Average (56.9%)</td>
<td></td>
</tr>
</tbody>
</table>

F. Oscillatory activity

Expected oscillatory features of a normal eyes open resting-state EEG in the central brain region is a pattern of alpha band activity and beta activity from occipital and frontal regions, respectively. These patterns match the collected data. Across subjects, increased levels of theta and alpha band activity are observed during non-control trials (MT, VIm, VInt). Theta and alpha activity has prior been correlated to activity in the auditory cortex, in regards processing speech and listening to music [37], [38]. Theta activity in EEG is also thought to represent bottom up processing doing working memory load in regards to auditory activity [39], [40].

G. Participant fatigue

Participants reported their ‘ability to concentrate’ level in questionnaires provided before and after the tasks, on a spectrum from 1 to 7, in which 7 is their highest concentration level and 1, the lowest. Participants were told to report their current concentration level prior to the tasks, and once more after the tasks, as well as reporting the perceived concentration level during the tasks. The measure of concentration ability is used as a tool to determine development of fatigue due to mental tasks and to gain an understanding of the cognitive taxation by the provided VIm tasks.

Based on the questionnaire response, a measure of the fatigue development among the participants before, during, and after the trials are deduced and illustrated in the raincloud plot shown in Fig.7. It can be seen that, overall, the participants reported a mean concentration ability 4.63 out of 7 (std: 0.78) prior to performing tasks, with a reported decrease in concentration ability during (mean: 4.5, std: 1.22) and after tasks (mean: 4.25, std: 1.3). A difference is observed when separating concentration level developments in gender (see Fig. 7). Male participants reported a slightly lower concentration level prior to tasks (mean: 4.5, std: 0.81), followed by a temporary increase during tasks (mean: 4.6, std: 1.36) and returned approximately to the same level after completing the tasks (mean: 4.5, std: 1.43). In comparison, female participants reported a higher level prior to beginning tasks (mean: 4.83, std: 0.69), with a decrease during tasks (mean: 4.33, std: 0.94) and further decrease after having completed all the tasks (mean: 3.83, std: 0.9). Although female participants, in general, showed higher accuracy rates for the different tasks, there is no individual correlation between development in concentration levels and classification rates. No correlation between age and concentration levels are found either. All trials are completed during afternoon and evening hours. There is no control to exclude natural fatigue, caused by tiredness during late hours of the day. There might be an indication that female subjects are more subject to mental fatigue from EEG-BCI VIm related cognitive load. However, the study population is too limited to draw any such conclusions.

Overall, the reported decrease in concentration levels and fatigue levels of participants are lower than expected, based on the usually high levels of fatigue reported in BCI studies making use of SSVEP and MI based paradigms.

IV. Discussion

In the presented work, we have used a SVM classification on CSP extracted features from EEG data corresponding to VIm. A cross-validation accuracy of above 95% is shown among individual participants and close to 89% across average of the 17 participants compared to other studies showing accuracy around 75% classifying VInt [19]. These accuracy scores are quite impressive considering no pre-processing was performed to prime the data; no artifact removal, frequency filtration or noise reduction. However, it is noteworthy that the weight matrix, W, is determined on subject basis for classification. This allows the model to adjust for potential variance of impedance for each electrode.

To the best of our knowledge, this is the first study including more than 10 participants for analyzing VIm, as well as the first study to analyze performance in both males and females.

VIm has shown to be a paradigm that has a high potential classification rate, with computational methods that require low level of processing power. In addition, VIm has shown to be easy to instruct to participants, and participants are able to perform VIm tasks in lengthy periods with continuous performance. Participants did report low levels of mental fatigue increase. None of the participants had received any prior training, besides a single test trial of 1 minute, without any feedback. All in all, these elements argue for recommending using VIm as a potent paradigm in the future EEG based BCIs.

VInt tend to perform better as a potential paradigm for classification than VIm. It is plausible that this task yields sources being
investigating spatial and temporal capacities of VIm classification, for its potential in advanced BCI systems. We propose studies for high accuracy, non-motor imagery based, multi function control will reveal whether this method for VIm will be useful for multi-
Further studies testing multiple different sounds, words and speech setup, but with issues regarding high levels of noise and artefact. also interesting due to EEGs nature, as a relatively easy and portable cases for the technology. More accuracy with EEG based systems are in developing new BCI tools and potentially enables more viable use-
underlines the potential of VIm/VInt in BCI systems with high levels of accuracy.
The high classification rates of VIm and VInt, with computational methods such as SVM that require low amount of processing power, underlines the potential of VIm/VInt in BCI systems with high levels of accuracy.
High levels of accuracy in EEG based BCI systems are interesting in developing new BCI tools and potentially enables more viable use-cases for the technology. More accuracy with EEG based systems are also interesting due to EEGs nature, as a relatively easy and portable setup, but with issues regarding high levels of noise and artefact. Further studies testing multiple different sounds, words and speech will reveal whether this method for VIm will be useful for multi-variable classification with likewise high levels of accuracy, allowing for high accuracy, non-motor imagery based, multi function control of BCI systems.
Further investigation in VIm based EEG analyses is interesting for its potential in advanced BCI systems. We propose studies investigating spatial and temporal capacities of VIm classification, as well as the role and use of conceptualization and semantics in VIm based BCI systems. A study with more than 32 channel EEG may show opportunities with the VIm paradigm in terms of capacity, as well as more detailed overview of activity throughout the cortex.
Other fields of interest for further study are the role of ‘inner voice’ and whether it can be observed in recordings with this methodology.

V. CONCLUSION
In conclusion, this study shows the viability of VIm as a highly classifiable EEG paradigm that has prospective use in EEG based BCI systems, and that these high levels of classification are possible with low levels of computational power. Further, the study compares EEG activity in producing vocal activity with imagery and intention. Results clearly indicate that VIm may yield better accuracy in classifying user commands compared to other currently used paradigms such as MI and more flexibility than other paradigms with high classification rates such as SSVEP based systems.

REFERENCES


