

Classifying motor imagery in presence of speech

Hayrettin Gürkök, Mannes Poel, Job Zwiers

Abstract— In the near future, brain-computer interface (BCI) applications for non-disabled users will require multimodal interaction and tolerance to dynamic environment. However, this conflicts with the highly sensitive recording techniques used for BCIs, such as electroencephalography (EEG). Advanced machine learning and signal processing techniques are required to decorrelate desired brain signals from the rest. This paper proposes a signal processing pipeline and two classification methods suitable for multiclass EEG analysis. The methods were tested in an experiment on separating left/right hand imagery in presence/absence of speech. The analyses showed that the presence of speech during motor imagery did not affect the classification accuracy significantly and regardless of the presence of speech, the proposed methods were able to separate left and right hand imagery with an accuracy of 60%. The best overall accuracy achieved for the 5-class separation of all the tasks was 47% and both proposed methods performed equally well. In addition, the analysis of event-related spectral power changes revealed characteristics related to motor imagery and speech.

I. INTRODUCTION

A brain-computer interface (BCI) facilitates direct communication from brain to computer without the need for any movement or other input modalities. This can be accomplished, for instance, by processing the electrical activity of the brain, using electroencephalography (EEG). There are different recording paradigms for BCIs such as sensorimotor rhythms (SMR) and event-related potentials (ERP). A good survey of BCI recording methods, paradigms and systems can be found in [1]. Until recently, the majority of the research done on BCIs aimed at developing assistive tools for disabled people. Nowadays, BCI applications for healthy users are also being considered. These applications usually allow users to control games or virtual environments (VEs) [2].

In VEs, especially in games, users show some reactions during play such as body movements which can provide the player a stronger affective experience [3]. They type or talk to the system or to the other players for mandatory or voluntary interaction. They continuously receive audio or visual feedback from the application. This sort of dynamic, multimodal interaction can not be tolerated to a great extent by BCIs. One of the reasons for this conflict is that the BCI recording systems are sensitive, so they pick up muscular activity along with the neurological activity. This forces the experimenters to instruct their subjects not to talk, move or even blink too much. Such muscular activity is usually considered as noise and during signal processing those noisy intervals are often removed. However, together with the noise, valuable activity information accompanying and representing the affective or

cognitive state of the subject is also lost. Another reason is the possible involvement of common brain areas in BCI related activities and other activities such as speaking and active listening. To tackle the challenges originating from these two reasons, advanced machine learning and signal processing techniques are required to decorrelate desired brain signals from the rest [4].

This work investigates how well speech can be used together with motor imagery (with respect to SMR) in a possible multimodal interface. For this purpose, three goals were set:

- *G1*: To determine how the classification of left and right hand motor imagery is affected when speech is present.
- *G2*: To see how well a system can classify left and right hand motor imagery regardless of presence of speech.
- *G3*: To assess the overall performance of a system to classify left and right hand motor imagery, the speech, and left and right hand motor imagery containing speech.

As no similar work was found in known literature, in this study, data was collected through a new experiment where the subjects performed left and right hand motor imagery while remaining silent and also while speaking. A signal processing chain and two classification methods were developed. The offline analyses revealed that the presence of speech during motor imagery did not affect the classification accuracy significantly and regardless of the presence of speech, the proposed methods were able to separate left and right hand imagery with an accuracy of 60%. The best overall accuracy achieved for the 5-class separation of all the tasks was 47% and both proposed methods were comparable in performance.

The rest of this paper is organised as follows. In Section 2 some related work is presented. The experimental setup and methods are described in Section 3. The results of the experiments are provided in Section 4. Section 5 summarises the experimental results and points to future research directions.

II. RELATED WORK

A. Realistic BCI settings

Recently there have been some studies investigating the effect of real-life conditions on BCI. Tangermann, et al. [4] reported on the interaction of subjects with a real pinball machine controlled by motor imagery. Compared to restricted lab settings the pinball machine provided rich and complex feedback, acoustic and visual distracters, and a challenging behavioural task. They showed that it was also possible to control the machine without user training.

Solovey, et al. [5] identified some considerations and provided guidelines for using fNIRS in realistic HCI settings.

The authors are with the Human Media Interaction Group, University of Twente, P.O. Box 217, 7500 AE, Enschede, the Netherlands (e-mails: {h.gurkok,m.poel,j.zwiers}@cs.utwente.nl).

They examined whether typical human behaviour (e.g., head and facial movements) or computer interaction (e.g., use of keyboard and mouse) interfere with fNIRS measurements. They stated that, providing the interference is corrected according to the guidelines, fNIRS can be used in realistic experiment environments.

Lotte, et al. [6] investigated the feasibility of using an EEG system based on P300 signals with a moving subject. They found that it was possible to detect the P300 signal while sitting, standing or walking.

An example BCI application for multimodal interaction was developed by Mühl, et al. [7] as a multi-paradigm game using EEG. In this game, called Bacteria Hunt, the player controls an amoeba which is trying to eat the fleeing bacteria. Movement direction is controlled by the keyboard and is also influenced by relative alpha power. When the amoeba is on a bacterium, eating is triggered by a flickering circle which stimulates SSVEP [1].

B. Motor imagery

Motor imagery can change the neuronal activity of the brain, especially in the primary sensorimotor area. Sensorimotor rhythms (SMR) are important phenomena for BCI that occur over the somatosensory cortices and include the μ (8-12 Hz) and the central β (13-28 Hz) rhythms [8]. SMRs increase at rest or in idling state [9] and decrease with movement, preparation for movement, and also with motor imagery [10]. The increase is called event-related synchronisation (ERS) and the decrease is called event-related desynchronisation (ERD).

The possibility of discriminating between left and right hand imagination based on ERD/ERS has been shown before. In the Wadsworth sensorimotor rhythm-based BCI system [11], people learn to manipulate their μ and β rhythm amplitudes which are then translated into cursor movements by a linear model. The recorded EEG signals are processed using Laplacian and common average spatial filters. Then FFT-based or matched-filter spectral analysis is carried out. The prediction weights are optimised using regression and the targets were made equally accessible by a form of normalisation of the resulting control signals. Similarly the Graz BCI system [12] estimates the power in the 5-35 Hz band using various methods such as the adaptive autoregressive model (AAR), common spatial patterns (CSP, explained in more detail in §III-D.1) and the hidden Markov model (HMM). The first two methods employ linear classification (LDA) while the third selects the maximum best path probability as the classifier. The system can also select subject specific frequency bands using distinction sensitive learning vector quantization (DSLQ).

III. METHODOLOGY

This section first provides statistics on subjects, then describes the experimental setup and the procedure. After that EEG recording and preprocessing steps are explained. Finally the filtering technique and the two classification methods are detailed.

A. Subjects

Eight subjects (three female) took part in the study. They had an average age of 25.75 (SD=1.04), ranging from 24 to 27 years. All were right-handed. The distribution of their native languages was as follows: one Chinese, three Turkish and four Dutch. Three subjects had previous experience with EEG, the rest were inexperienced. None of the participants had any known neurological disorders or other significant health problems. Informed consent was obtained from all subjects and the student volunteers were paid for their participation.

B. Experimental tasks and procedure

Five experimental tasks were performed during EEG recording. Before the experiment the subjects received a briefing and a number of training trials until they felt confident enough that they could perform all the tasks. In order to isolate only the speech and motor imagery, the subjects were asked not to talk or move unless instructed to do so. Subjects' voices were also recorded.

The experimental tasks performed are described in Table I. For the tasks including motor imagery (L, R, LS and RS), subjects were asked to imagine moving their left or right hand kinesthetically [8]. For the tasks including speech (LS, RS, S) they were asked to talk continuously in their mother tongue. To ensure the existence and naturalness of speech, they were shown a picture of a different object for each task. For this purpose 150 pictures, detailed and general enough to stimulate speech, were chosen from the Amsterdam Library of Object Images (full color, quarter resolution, viewing direction) [13]. The pictures and their order of presentation were the same for all subjects. Subjects could describe the pictures, express their opinion or say anything related to the pictures. Even if they had no idea about the picture they were requested to say so to ensure presence of speech. For the tasks including both speech and motor imagery (LS and RS) they were asked to start performing both subtasks simultaneously and avoid executing them sequentially.

TABLE I
TASKS DURING EXPERIMENTS

L	Imagine you are moving your left hand
R	Imagine you are moving your right hand
LS	Imagine you are moving your left hand at the same time speak about the picture shown earlier
RS	Imagine you are moving your right hand at the same time speak about the picture shown earlier
S	Talk freely about the picture shown earlier

The experiment consisted of five sessions with breaks of unrestricted duration in between. Each session contained ten trials per task shuffled randomly which meant that each task was repeated fifty times during the experiment. So there were fifty trials per session and a total of 250 trials during the experiment.

Each trial lasted eight seconds. During the first three seconds of a trial for the non-speech tasks (L and R) a

pause symbol was displayed on the screen while for the other tasks (LS, RS, S) a picture stimulus was shown to facilitate the speech. Throughout the remaining five seconds a cue image was displayed so that the subject starts to do the task instructed. The structure and timing for trials is shown in Figure 1.

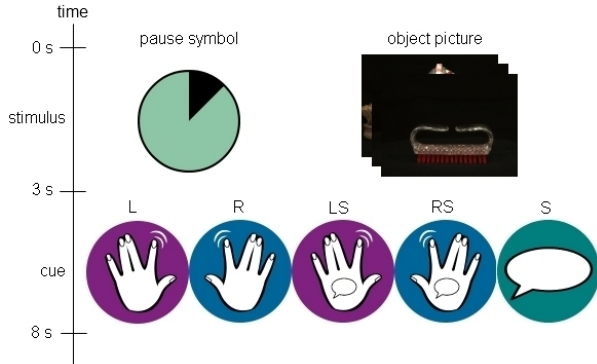


Fig. 1. Structure and timing for a trial

C. EEG recording and preprocessing

EEG activity was recorded from 32 electrode sites according to the international 10-20 system [14] using an electrode cap (BioSemi Active electrodes and headcap) placed on the subject's head. The continuous EEG signals were recorded and digitised at a sampling rate of 512 Hz using BioSemi ActiveTwo system. No further processing was carried out in the hardware. Signal preprocessing was carried out using the EEGLAB toolbox [15].

The steps of EEG processing and analysis are depicted in Figure 2. The digitised EEG signals were first converted to common average reference (CAR) to increase the signal-to-noise ratio (SNR) and reduce drifts [16]. Next a bandpass filter of 8-28 Hz was applied to get rid of the lower and higher frequencies which were not of interest. The preprocessed signal was then broken into trial-wise epochs. Each trial lasted eight seconds; the first three seconds contained the cue presentation and the remaining five seconds the actual motor imagery task. An epoch consisted of the last five seconds of a trial. As there were 250 trials there were also 250 epochs, 50 per task.

To properly estimate the classification accuracy, epochs of each task were divided into training and testing tests. This training/testing procedure was repeated 10 times using different random partitions for training and testing sets (i.e. 10-fold cross-validation) [17]. To ensure the stability, the final classification performance was reached by repeating the 10-fold cross-validation three times and averaging the cross-validation performances (i.e. 3x repeated 10-fold cross-validation).

D. EEG analysis

EEG signals recorded from the scalp are often so weak that they get distorted by stronger signals, such as those generated by muscle movements. Therefore, especially in single-trial

experiments, spatial filtering is helpful to improve the SNR of the EEG. Among familiar spatial filters such as bipolar, CAR and Laplacian, common spatial patterns (CSP) filters are shown to have the most discriminative power for motor imagery [18]. In the case of principal component analysis (PCA) the spatial filters are designed so that each temporal EEG sequence extracted contains a maximum portion of the combined temporal variance. However, in CSP the filters are designed so that they can extract temporal sequences of maximum variance from one class and those of minimum variance from a second class [19].

Each goal listed in §I requires classification of different number of classes. $G1$ compares L vs. R and LS vs. RS classifications, $G2$ targets all left (L and LS) vs. all right (R and RS) classification while $G3$ investigates separation of all the five tasks (i.e. L, R, LS, RS and S). Therefore, $G1$ and $G2$ demand dichotomous classifications while $G3$ a polychotomous one.

As stated above, CSP filters are useful in discriminating exactly two populations of EEG patterns but in case of $G3$ there are five classes. Thus, firstly, the details of constructing the CSP filter for a class pair will be described and then two generic multiclass separation algorithms will be presented.

1) *CSP filtering*: The CSP filtering is implemented as described in [20]. The preprocessed (as described in Section III-C) EEG data of every training epoch is represented as an $N \times T$ matrix X where N is the number of channels (i.e. recording electrodes) and T is the number of samples per channel. First the normalised spatial covariance of epoch X is obtained from:

$$C(X) = \frac{XX'}{\text{trace}(XX')} \quad (1)$$

where $\text{trace}(x)$ is the sum of the diagonal elements of x . The normalisation of C with respect to the trace is done to eliminate magnitude variations in the EEG that exist between individuals. The diagonal elements of C now represent a measure of the fractional variance (i.e. fraction of the total power) of each EEG channel and the off-diagonal elements the fractional covariance.

The rest of this subsection describes how the CSP filter is computed for a class (i.e. task) pair (a, b) where $a, b \in Q$ and

$$Q = \{L, R, LS, RS, S\} \quad (2)$$

The population covariance matrices R_a and R_b are computed by averaging the individual covariance matrices of all training epochs for classes a and b respectively. A composite covariance matrix R_c is obtained from:

$$R_c = R_a + R_b \quad (3)$$

R_c is then factored into its matrices of eigenvectors (U_c) and eigenvalues (λ_c) as:

$$R_c = U_c \lambda_c U_c' \quad (4)$$

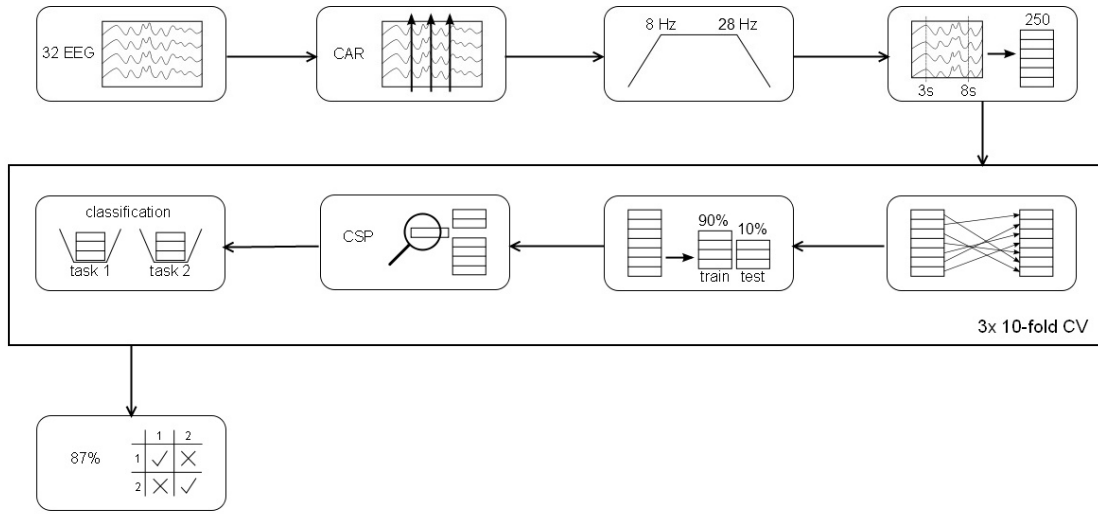


Fig. 2. Steps of EEG processing and analysis: channel selection, re-referencing, bandpass filtering, epoching, shuffling, partitioning, CSP filtering, classification and evaluation of results.

Throughout this subsection, all the eigenvalues are sorted in descending order and the eigenvectors accordingly. After that, the whitening transformation matrix P is formed by:

$$P = \sqrt{\lambda_c^{-1}} U'_c \quad (5)$$

which equalises the variances in the space spanned by U_c so that all eigenvalues of PR_cP' are equal to one. If R_c is ill-conditioned then P is computed using only the most significant eigenvalues and vectors. In this case P will be of size $M \times N$.

As shown in [21], if R_a and R_b are transformed as:

$$S_a = PR_aP' \text{ and } S_b = PR_bP' \quad (6)$$

then S_a and S_b share common eigenvectors and the corresponding eigenvalues for the two matrices will always sum up to 1. Therefore if S_a is factored as:

$$S_a = B\psi_aB' \quad (7)$$

then S_b can be factored as $S_b = B\psi_bB'$ and the following will hold:

$$\psi_a + \psi_b = I \quad (8)$$

Since the sum of the corresponding eigenvalues is always one (by (8)), the eigenvector with the largest eigenvalue for S_a has the smallest eigenvalue for S_b and vice versa. Therefore the common eigenvector B yields a decomposition which is optimal for separating the variances in the two EEG classes. Now, a projection matrix is formed as follows:

$$W = (P'B)' \quad (9)$$

where the columns of W^{-1} are common spatial patterns and can be considered as time-invariant EEG source distribution vectors.

The projection matrices are computed by repeating this procedure for all possible class pairs.

2) *Multiclass separation*: In order to determine the features, a testing epoch x is first filtered separately for each class pair (i, j) as:

$$Z(x)_{(i,j)} = W_{(i,j)} \cdot x \quad (10)$$

where $i, j \in Q$, $i \neq j$ and $W_{(i,j)}$ is the projection matrix calculated for pair (i, j) as in (9).

Z projects, along its rows, the variance of i in x which is largest along the first row and increasingly less along the subsequent rows. Therefore the feature vectors $f(x)_{(i,j)}$ are formed by taking the variances of the first and last m rows of $Z(x)_{(i,j)}$, which are the most suitable signals for separating the two classes i and j . m is fixed to 3 which was found to be the optimal value for this particular study in our preliminary tests. If the number of rows that Z contains is smaller than $2m$, then the maximum even number of rows is used.

As described above, the feature vectors (f) depend on the CSP filters (W) computed exclusively for two classes. Therefore the five classes do not reside in the same space but pairwise in different spaces which makes them incomparable. For this reason, we first employ a pairwise separation on each class pair. We find the similarity of x to a class i against another class j as: as follows:

$$s(x)_{ij} = \frac{\|f(x)_{(i,j)} - \mu_j\| - \|f(x)_{(i,j)} - \mu_i\|}{\|f(x)_{(i,j)} - \mu_j\| + \|f(x)_{(i,j)} - \mu_i\|} \quad (11)$$

where μ_i and μ_j are the mean feature vectors for classes i and j respectively and $\|v\|$ is the Euclidean length of a vector v . The denominator ensures that the final value is normalised within the range $[-1, 1]$. According to (11), the more similar the epoch is to class i , the closer s_{ij} is to 1. Also note that:

$$s(x)_{ij} = -s(x)_{ji} \quad (12)$$

Then, we follow an adapted soft-voting approach [22]. The idea behind the original soft-voting technique is to construct two-class decision boundaries independently between every pair of classes and use these boundaries to assign an unknown observation to one of its two respective classes. The individual class that receives the most such assignments over the decisions is taken as the predicted class for the observation. In our study, instead of crisp assignments, we use the similarity values computed as in (11). The predicted class for epoch x is the one that achieves the maximum pairwise similarity score:

$$label(x)_Q = \arg \max_{i \in Q} \left\{ \sum_{j \in Q} s_{ij}(x) \right\} \quad (13)$$

3) *Separation by decision trees*: Another approach to classifying a testing epoch x is by feeding it to a decision tree [17]. One could hypothesize that speech-including classes (LS, RS, S) would share common features which are different from those of non-speech classes (L and R) due to the difference in the power they contain. Similarly, the speech-only class (S) could possess different characteristics than speech-including motor imagery classes (LS and RS). Thus, dividing the classification problem hierarchically into subgroups would earn performance increment in comparison to multiclass separation described in §III-D.2. For this purpose, two decision trees, DT1 and DT2, are constructed as seen in Figure 3. Decision nodes use $label(x)_{Q_i}$ values computed by (13) with respect to the training sets (Q_i) formed as explained in Table II.

Using the decision trees, x is first marked for containing speech. If it does not, it is further classified as L or R. If it does, it is classified as LS, RS or S in case of DT1. If DT2 is used, it is fed to a node that decides whether x is

TABLE II
TRAINING SETS FORMED FOR DECISION TREES
 $MI=\{L \cup R\}$, $MISp=\{LS \cup RS\}$, $MISp+S=\{MISp \cup S\}$

Training set	Contained tasks
Q1	MI, MISp+S
Q2	L, R
Q3	LS, RS, S
Q4	MISp, S
Q5	LS, RS

pure speech (S) or speech-including motor imagery. If it is a speech-including motor imagery epoch, then it is classified as LS or RS .

Note that the method described in §III-D.2 can be considered as a decision tree consisting of a single decision node which uses $label(x)_Q$ as the attribute to classify the epoch x . Also note that in case of dichotomous classification, executions of both separation methods reduce to the same procedure.

IV. RESULTS

A. Classification of motor imagery with respect to speech

As shown in Table III, the classification accuracy achieved for left and right hand motor imagery was 60%. If the motor imagery contained speech, as seen in Table IV, performance dropped at all metrics except for the recall of right hand imagery. However, all the changes were insignificant. Throughout this section, the significance was assessed by two-tailed paired t-test ($p < 0.01$).

B. Classification of motor imagery regardless of speech

The performance for classifying motor imagery regardless of presence of speech is given in Table V. No significant difference was observed in comparison to the cases where

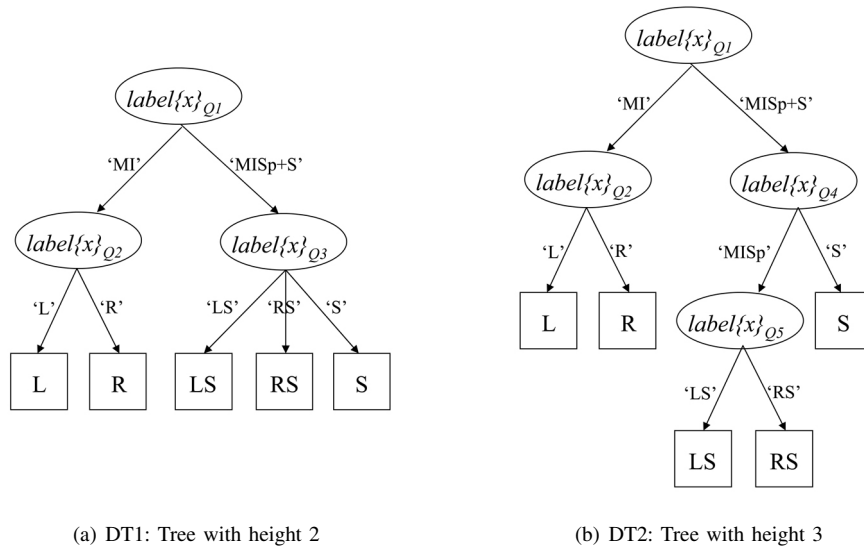


Fig. 3. The decision trees. The ellipses are the decision nodes where the filtering and classification are performed. The squares are the leaf nodes where the evaluation and labeling take place. MI is the set of all non-speech trials, MISp+S is the set of all trials containing speech and MISp is the set of motor imagery trials containing speech.

speech was always absent or present (Tables III and IV respectively).

TABLE III

AVERAGE CLASSIFICATION ACCURACIES AND THE RECALL AND PRECISION RATES FOR LEFT AND RIGHT HAND MOTOR IMAGERY. STANDARD DEVIATIONS ARE IN PARENTHESES.

Accuracy	L		R	
	Recall	Prec	Recall	Prec
0.60 (0.15)	0.58 (0.26)	0.58 (0.17)	0.62 (0.20)	0.61 (0.17)

TABLE IV

AVERAGE CLASSIFICATION ACCURACIES AND THE RECALL AND PRECISION RATES FOR LEFT AND RIGHT HAND MOTOR IMAGERY WITH SPEECH. STANDARD DEVIATIONS ARE IN PARENTHESES.

Accuracy	LS		RS	
	Recall	Prec	Recall	Prec
0.57 (0.14)	0.49 (0.23)	0.56 (0.18)	0.65 (0.18)	0.56 (0.15)

TABLE V

AVERAGE CLASSIFICATION ACCURACIES AND THE RECALL AND PRECISION RATES FOR LEFT AND RIGHT HAND MOTOR IMAGERY REGARDLESS OF PRESENCE OF SPEECH. STANDARD DEVIATIONS ARE IN PARENTHESES.

Accuracy	L \cup LS		R \cup RS	
	Recall	Prec	Recall	Prec
0.60 (0.13)	0.56 (0.24)	0.59 (0.14)	0.64 (0.18)	0.61 (0.14)

C. Overall classification performances

Table VI displays the average (over subjects) classification accuracy together with the precision and recall rates achieved by the two methods: multiclass CSP separation (Multi-CSP) and decision trees (DT1 and DT2). The values in parentheses are the standard deviation values.

The improvement expected through the decision trees was not observed in the final performance. Multi-CSP not only yielded a performance close to those of the decision trees but also enhanced them significantly in RS recall (with respect to DT1 and DT2) and S recall (with respect to DT1).

However, when the performances of the decision trees are considered per level, the hypotheses suggested in §III-D.3 still hold. We can list the possible findings as follows:

- 1) The mean accuracy obtained during the very first split (for $Q1$) was 0.87 (SD=0.04) which implies that speech-including classes share features different from those of non-speech classes.
- 2) The significant improvement obtained by DT2 in comparison to DT1 for S recall (in $Q4$ classification, compared to $Q3$ classification) implies that speech-including motor imagery classes also have different characteristics than the speech-only class.

- 3) The relatively low recall and precision rates during $Q2$ and $Q5$ classifications could suggest that left and right motor imagery trials contain very common features, both in presence or absence of speech, hence they are not easily discriminable.

A more conservative frequency filtering for the 8-12 Hz band was also tested so as to consider the μ -rhythm only. No significant difference was observed in any metric.

D. Speech and motor imagery related dynamics

Event-related spectral perturbation (ERSP) is the event-related shift in the power spectrum at a certain time and frequency [23]. Figure 4 displays the ERSP change per task for the right sensorimotor channel, C4, for saving the space. Similar plots were obtained for the sensorimotor channels C3 and Cz. The time ranges from 2 seconds before the cueing (which occurs at time 0, indicated with a dashed line) to 2 seconds after the trial ends (i.e. 9 seconds in total) and the frequency ranges from 8 Hz to 32 Hz. Only the significant power changes with respect to the baseline (2 seconds before the cueing) according to the bootstrap level of 0.05 are coloured according to the colour bar at the bottom, the rest are in green. The two series depicted under the ERSP plot are the low and high mean dB values, relative to the baseline, at each time point and on the left side lies the baseline mean power spectrum.

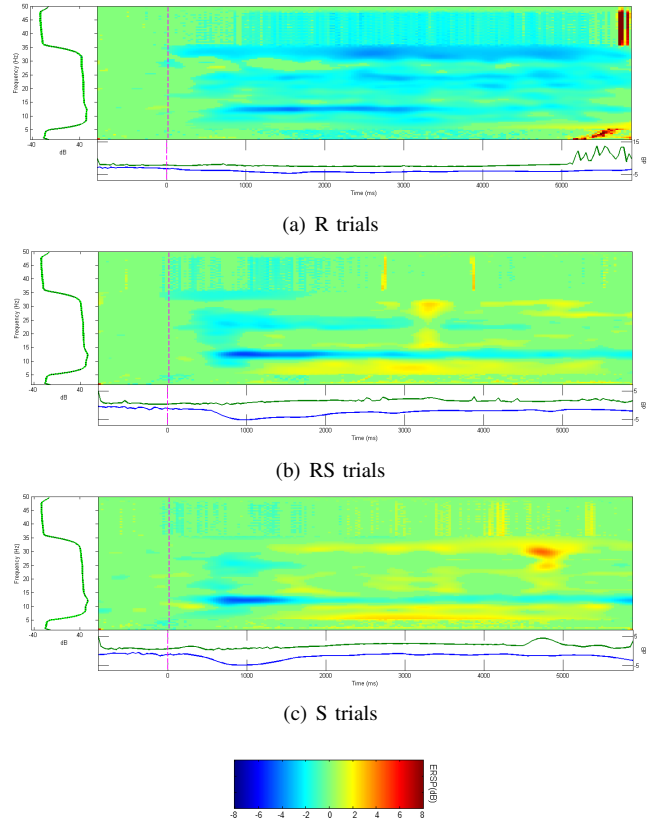


Fig. 4. ERSP plots for all trials averaged over all subjects for the channel location C4. Similar plots were obtained for the sensorimotor channels C3 and Cz. Colours are visible in online version.

TABLE VI

CLASSIFICATION ACCURACIES TOGETHER WITH THE RECALL AND PRECISION RATES AVERAGED OVER SUBJECTS. STANDARD DEVIATIONS ARE IN PARENTHESES. THE HIGHEST PERFORMANCE PER COLOUMN IS INDICATED IN BOLD.

Method	Accuracy	L		R		LS		RS		S	
		Recall	Prec	Recall	Prec	Recall	Prec	Recall	Prec	Recall	Prec
Multi-CSP	0.47 (0.11)	0.52 (0.30)	0.50 (0.16)	0.47 (0.24)	0.52 (0.12)	0.32 (0.18)	0.42 (0.17)	0.46 (0.24)	0.39 (0.17)	0.59 (0.18)	0.56 (0.16)
DT1	0.45 (0.08)	0.55 (0.27)	0.43 (0.13)	0.50 (0.24)	0.49 (0.14)	0.31 (0.21)	0.35 (0.16)	0.37 (0.17)	0.40 (0.14)	0.51 (0.14)	0.55 (0.18)
DT2	0.44 (0.08)	0.53 (0.29)	0.44 (0.12)	0.49 (0.25)	0.47 (0.15)	0.29 (0.18)	0.38 (0.16)	0.31 (0.17)	0.37 (0.10)	0.60 (0.09)	0.52 (0.16)

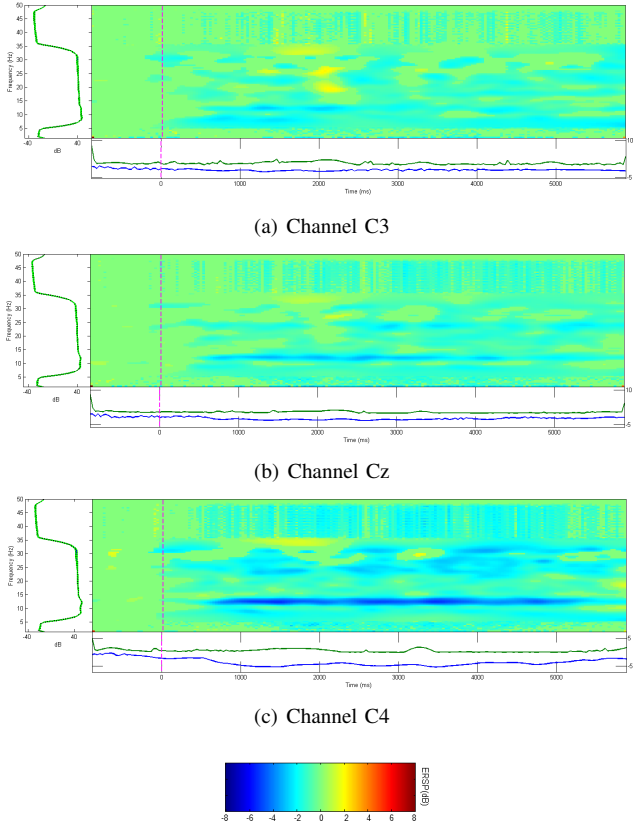


Fig. 5. ERSP plots for the left hand imagination trials averaged over all subjects for the sensorimotor channel locations C3, Cz and C4. Similar plots were obtained for the remaining four tasks as well. Colours are visible in online version.

The purpose of inspecting the ERSP plots is to verify the list of findings mentioned in §IV-C. Figure 4(a) (representing non-speech trials) shows the characteristic bilateral ERD during motor imagery by the power decrease in alpha (8-12 Hz) and central beta (13-28 Hz) frequencies throughout the whole trial. This major pattern in both non-speech classes is in compliance with Item 3. When speech is involved, in Figure 4(b), the power decrease in the lower beta is less visible and there are power increases in higher central beta frequencies. This change in the pattern conforms with Item 1. In speech-only trials (Figure 4(c)) there is a more intense power increase in beta frequency peaking towards the end of

the trial. Finally, this difference with respect to the speech-including motor imagery trials is in accordance with Items 2 and 3.

The ERSP plots also revealed that, in all tasks, the power decrease is more intense in channel C4 than in channels C3 and Cz. Figure 5 displays the case for the left hand imagination task.

V. CONCLUSIONS

Throughout this paper we sought to understand how well motor imagery can be classified in the presence of speech. We devised an experimental setup where the subjects imagined moving their left and right hands and sometimes talked simultaneously about stimuli that they were shown. We proposed a signal processing pipeline and two classification methods. The first classification method is the multiclass separation based on the CSP filtering technique. The latter is the decision tree approach making use of the former method as the decision mechanism. Both methods yielded similar performance in most of the metrics evaluated, including the overall classification accuracy.

In terms of the goals set in the beginning of this paper, we found that presence of speech during motor imagery did not affect the classification accuracy significantly ($G1$). Regardless of the presence of speech, the proposed method was able to separate left and right hand imagery with an accuracy of 60% ($G2$). The best overall classification accuracy achieved was 47%, which was significantly higher than the five-class chance level of 20% ($G3$).

Observing decision tree classifications in node-level provided us with some insights about the characteristics of speech and motor imagery signals. We found that speech-including signals and non-speech signals possess characteristics very different from each other. Similarly, speech-including motor imagery tasks yielded features dissimilar to those of the speech-only task. However, left and right imaginary tasks were not easily separable, meaning that their signals contain similar patterns.

We examined the event-related power changes in the brain with respect to speech and motor imagery, averaged over all subjects. The plots we obtained were in-line with the analysis of the decision tree classification. During motor imagery of either hand, a bilateral desynchronisation in μ

and central β rhythms was observed, biased towards the right hemisphere. The desynchronisation in the β rhythm faded when the subjects spoke. When there was no motor imagery but only speech, μ rhythm desynchronisation also faded and significant β synchronisation was observed.

In this work all the subjects were right-handed so it might be useful to validate the findings on a population counterbalanced on handedness. As indicators of motor imagery, we used SMR. Other event-related features, such as the lateralised readiness potential (LRP) [24], may also be employed.

This offline study allowed us to evaluate statistically the classification performance of the methods we proposed. A possible next step would be to employ these methods in an interactive system, such as a speech and brain commanded interface, in order to assess its usability in realistic settings.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the support of the BrainGain Smart Mix Programme of the Netherlands Ministry of Economic Affairs and the Netherlands Ministry of Education, Culture and Science. The icons used in Figure 2 are courtesy of Marijn van Vliet and the cue images used during the experiment are courtesy of Corona L. Zschüsschen.

REFERENCES

- [1] A. Kübler and K.-R. Müller, *An Introduction to Brain-Computer Interfacing*, ser. Neural Information Processing. Cambridge, MA, USA: The MIT Press, 2007, pp. 1–25.
- [2] A. Nijholt, D. Tan, G. Pfurtscheller, C. Brunner, J. d. R. Millán, B. Allison, B. Graimann, F. Popescu, B. Blankertz, and K.-R. Müller, “Brain-computer interfacing for intelligent systems,” *IEEE Intelligent Systems*, vol. 23, no. 3, pp. 72–79, 2008.
- [3] N. Bianchi-Berthouze, W. W. Kim, and D. Patel, “Does body movement engage you more in digital game play? And why?” in *Proceedings of the 2nd International Conference on Affective Computing and Intelligent Interaction*. Berlin/Heidelberg, Germany: Springer-Verlag, 2007, pp. 102–113.
- [4] M. Tangermann, M. Krauledat, K. Grzeska, M. Sagebaum, B. Blankertz, C. Vidaurre, and K.-R. Müller, “Playing Pinball with non-invasive BCI,” in *Advances in Neural Information Processing Systems 21*. Cambridge, MA, USA: The MIT Press, 2008, pp. 1641–1648.
- [5] E. T. Solovey, A. Girouard, K. Chauncey, L. M. Hirshfield, A. Sasaroli, F. Zheng, S. Fantini, and R. J. Jacob, “Using fNIRS brain sensing in realistic HCI settings: Experiments and guidelines,” in *Proceedings of the 22nd Annual ACM Symposium on User Interface Software and Technology*. New York, NY, USA: ACM, 2009, pp. 157–166.
- [6] F. Lotte, J. Fujisawa, H. Touyama, R. Ito, H. Michitaka, and A. Lécuyer, “Towards ambulatory brain-computer interfaces: A pilot study with P300 signals,” in *Proceedings of the 5th Advances in Computer Entertainment Technology Conference*, 2009.
- [7] C. Mühl, H. Gürkök, D. Plass-Oude Bos, M. E. Thurlings, L. Scherffig, M. Duvinage, A. A. Elbakyan, S. Kang, M. Poel, and D. Heylen, “Bacteria Hunt: A multimodal, multiparadigm BCI game,” in *Proceedings of the 5th International Summer Workshop on Multimodal Interfaces*, 2010.
- [8] C. Neuper, R. Scherer, M. Reiner, and G. Pfurtscheller, “Imagery of motor actions: Differential effects of kinesthetic and visual-motor mode of imagery in single-trial EEG,” *Cognitive Brain Research*, vol. 25, no. 3, pp. 668–677, 2005.
- [9] G. Pfurtscheller, “Event-related synchronization (ERS): An electrophysiological correlate of cortical areas at rest,” *Electroencephalography and Clinical Neurophysiology*, vol. 83, no. 1, pp. 62 – 69, 1992.
- [10] G. Pfurtscheller and C. Neuper, “Motor imagery activates primary sensorimotor area in humans,” *Neuroscience Letters*, vol. 239, no. 2-3, pp. 65–68, 1997.
- [11] D. J. McFarland, D. J. Krusienski, and J. R. Wolpaw, “Brain-computer interface signal processing at the Wadsworth Center: Mu and sensorimotor beta rhythms,” *Progress in Brain Research*, vol. 159, pp. 411–419, 2006.
- [12] G. Pfurtscheller and C. Neuper, “Motor imagery and direct brain-computer communication,” *Proceedings of the IEEE*, vol. 89, no. 7, pp. 1123–1134, 2001.
- [13] J. M. Geusebroek, G. J. Burghouts, and A. W. M. Smeulders, “The Amsterdam Library of Object Images,” *International Journal of Computer Vision*, vol. 61, no. 1, pp. 103–112, 2005.
- [14] H. Jasper, “Report of the committee on methods of clinical examination in electroencephalography,” *Electroencephalography and Clinical Neurophysiology*, vol. 10, pp. 370–375, 1958.
- [15] A. Delorme and S. Makeig, “EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis,” *Journal of Neuroscience Methods*, vol. 134, pp. 9–21, 2004.
- [16] D. J. McFarland, L. M. McCane, S. V. David, and J. R. Wolpaw, “Spatial filter selection for EEG-based communication,” *Electroencephalography and Clinical Neurophysiology*, vol. 103, no. 3, pp. 386–394, 1997.
- [17] E. Alpaydin, *Introduction to Machine Learning*. Cambridge, MA, USA: The MIT Press, 2004.
- [18] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, and K.-R. Müller, “Optimizing spatial filters for robust EEG single-trial analysis,” *IEEE Signal Processing Magazine*, vol. 25, no. 1, pp. 41–56, 2008.
- [19] Z. J. Koles, “The quantitative extraction and topographic mapping of the abnormal components in the clinical EEG,” *Electroencephalography and Clinical Neurophysiology*, vol. 79, no. 6, pp. 440–447, 1991.
- [20] Z. J. Koles, L. J. C., and F.-H. P., “Spatial patterns in the background EEG underlying mental disease in man,” *Electroencephalography and Clinical Neurophysiology*, vol. 91, no. 5, pp. 319–328, 1994.
- [21] K. Fukunaga, *Introduction to statistical pattern recognition (2nd ed.)*. San Diego, CA, USA: Academic Press Professional, Inc., 1990.
- [22] J. H. Friedman, “Another approach to polychotomous classification,” Department of Statistics, Stanford University, Tech. Rep., 1996.
- [23] S. Makeig, “Auditory event-related dynamics of the EEG spectrum and effects of exposure to tones,” *Electroencephalography and Clinical Neurophysiology*, vol. 86, no. 4, pp. 283 – 293, 1993.
- [24] M. G. Coles, G. Gratton, and E. Donchin, “Detecting early communication: Using measures of movement-related potentials to illuminate human information processing,” *Biological Psychology*, vol. 26, pp. 69–89, 1988.