2019 IEEE 16th International Conference on Rehabilitation Robotics (ICORR)
Toronto, Canada, June 24-28, 2019

Decoding covert speech for intuitive control of brain-computer interfaces based on single-trial EEG: a feasibility study

L. Tøttrup, K. Leerskov, J. T. Hadsund, E. N. Kamavuako, *Member, IEEE,* R. L. Kæseler, and M. Jochumsen

Abstract—**For individuals with severe motor deficiencies, controlling external devices such as robotic arms or wheelchairs can be challenging, as many devices require some degree of motor control to be operated, e.g. when controlled using a joystick. A brain-computer interface (BCI) relies only on signals from the brain and may be used as a controller instead of muscles. Motor imagery (MI) has been used in many studies as a control signal for BCIs. However, MI may not be suitable for all control purposes, and several people cannot obtain BCI control with MI. In this study, the aim was to investigate the feasibility of decoding covert speech from single-trial EEG and compare and combine it with MI. In seven healthy subjects, EEG was recorded with twenty-five channels during six different actions: Speaking three words (both covert and overt speech), two arm movements (both motor imagery and execution), and one idle class. Temporal and spectral features were derived from the epochs and classified with a random forest classifier. The average classification accuracy was 67 ± 9 % and 75 ± 7 % for covert and overt speech, respectively; this was 5-10 % lower than the movement classification. The performance of the combined movement-speech decoder was 61 ± 9 % and 67 ± 7 % (covert and overt), but it is possible to have more classes available for control. The possibility of using covert speech for controlling a BCI was outlined; this is a step towards a multimodal BCI system for improved usability.**

I. Introduction

Diseases or injuries affecting the nervous system can result in impairment of the ability to control the body and produce movement [1], [2]. Conditions such as amyotrophic lateral sclerosis, cerebral palsy, stroke or spinal cord injury can lead to partial or total loss of muscular control [2]. Depending on the severity of the injury, different assistive technologies can be useful; several rely on some degree of muscular control such as robots controlled using keyboards and joysticks [3]. Individuals with severe disabilities are not able to control such devices. Brain-computer interfaces (BCIs) are often used as communication tools and for controlling neural prosthetics and other assistive technologies [4] and can be used by patients not able to control other assistive devices. The control of BCIs in general have improved with the advances in the data recording equipment and signal processing techniques, but successful use of BCIs is still heavily dependent on the users' ability to operate the BCI [5]. Control of BCIs may require weeks or months of training to acquire high accuracy for both healthy subjects and patients [6]–[9]. One approach to improve the performance of BCIs is to combine different signal modalities or control signals to form a hybrid BCI [10], [11]. BCIs are often controlled by motor imagery (MI), but it may be difficult for some users to imagine movements and associate these with actions of the device, and a non-negligible part of people are not able to use a MI-controlled BCI [12]. Therefore, it could be speculated that MI could be supplemented or replaced by other control signals that may be more intuitive for the user to control or easier to elicit.

Studies using EEG have shown that a combination of different mental tasks is feasible for BCI control [13], [14]. For example, in a four class BCI using a combination of abstract mental tasks (rotating Rubik's cube, mental counting, etc.) and MI tasks, an average accuracy of 62 % could be reached [13]. In a three class BCI using letter composing, arithmetic, and rotating Rubik's cube, average classification accuracies of 78 % and 72 % were obtained for able bodied subjects and tetraplegics, respectively [14]. However, for non-movement related actions, covert speech might be more intuitive as it enables a wide set of voluntary mental commands linguistically linked to specific actions in the use of an external device (e.g. a wheelchair driving forward by covertly saying "go", or a robotic arm grasping an object by covertly saying "grasp").

Previous research have used brain signals recorded with ECoG (Electro-corticography, implanted electrography) to predict consonants and vowels in imagined words [15]. In a review of covert speech decoding using ECoG, it is argued that the technology may successfully decode single words above chance level but not continuous speech [16]. Even though studies with ECoG shows promising results, the technology is invasive, and considering the current stage of the covert speech research, a non-invasive technology would be preferred. Several covert speech studies based on EEG have focused on comparing different vowels [17], vowels and consonants [18], or verbs and nouns [19] resulting in classification accuracies between 38 % (four classes) and 98 % (two classes).

Thus, the aim of this study is to investigate the feasibility of classifying speech (covert and overt) for controlling an external device based on single-trial EEG. In addition, speech classification is compared to and combined with classification of movements (imagination and execution). To the authors' knowledge, this is the first attempt at combining covert speech and MI for potential BCI control.

L. Tøttrup, (email: ltj@hst.aau.dk) K. Leerskov, (email: kkl@hst.aau.dk), J. T. Hadsund, R. L. Kæseler (rlk@hst.aau.dk), and M. Jochumsen are with SMI, Department of Health, Aalborg University, Denmark (phone: + 45 9940 3789; e-mail: mj@hst.aau.dk).

E. N. Kamavuako is with the Centre for Robotics Research, Department of Informatics, King's College London, United Kingdom (e-mail: ernest.kamavuako@kcl.ac.uk).

978-1-7281-2755-2/19/$31.00 ©2019 IEEE

II. METHODS

A. Subjects

Seven healthy right-handed subjects (five women; 24 ± 2 years old) participated. The subjects gave their written informed consent prior to participation. All procedures were approved by the local ethical committee (Region North Jutland, Denmark - N-20130081).

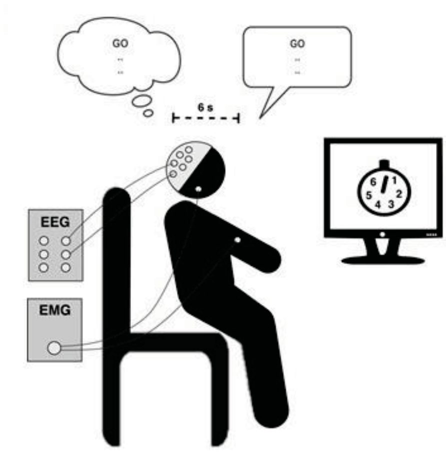
B. Data acquisition

25 passive Ag/AgCl sintered EEG ring electrodes were placed over the frontal, central, and parietal areas according to the international 10-20 system (AFz, Fz, F3-4, F7-8, FC1-2, FC5-6, FT9-10, Cz, C3-4, CP1-2, CP5-6, TP9-10, P3-4, and P7-8) with the impedance checked and kept below 10 kΩ during the experiment. The reference and ground electrodes were placed on the earlobes. Three Ag/AgCl surface Electromyography (EMG) electrodes (AMBU self-adhesive EMG electrodes) were placed below the jaw on the anterior belly of the digastric muscle and on the muscle belly of both biceps brachii. The EMG was used for synchronization of action onset in the EEG by identifying overt action onset from the electrodes recording ME of the biceps and overt speech from the digastric muscle. EEG and EMG were sampled with 500 Hz and converted with 32-bits accuracy (EEG amplifiers, Nuamps Express, Neuroscan).

C. Experimental procedure

The subjects were seated in a comfortable chair facing a computer screen displaying an analogue clock (**Figure 2**). The subjects were instructed to perform one of the following six actions: idling, speaking the control-related word “go” or “stop”, speaking a control-unrelated word: “Viborg” (a Danish city), flexing the right arm, or flexing the left arm. Every action was performed covertly (internal speaking or MI) and repeated overtly (external speaking or ME) after six seconds, corresponding to one clock cycle. The interval between repetitions of actions was 6-12 seconds (1-2 clock cycles) controlled by the subject (self-paced). Actions were in blocks of 4-8 repetitions (assigned randomly). A total of 14 blocks of randomly selected actions were performed followed by an eight-minute break. Idle activity was recorded for 30 seconds immediately before and after the 14 blocks of actions (**Figure 1**). This procedure was repeated until a total of 80 repetitions of each action were recorded. As covert speech/MI always occurred six seconds before overt speech/ME, the covert action onset could be identified, using the overt action onset. Due to time constraints for some participants, leading to incomplete measures, and epoch rejection of noisy epochs, an average of 62 ± 21 epochs remained for each overt action, and 61 ± 21 for each covert action, per subject.

Figure 2. Experimental setup. The subject was wearing EEG and EMG electrodes and faced a screen with a clock.

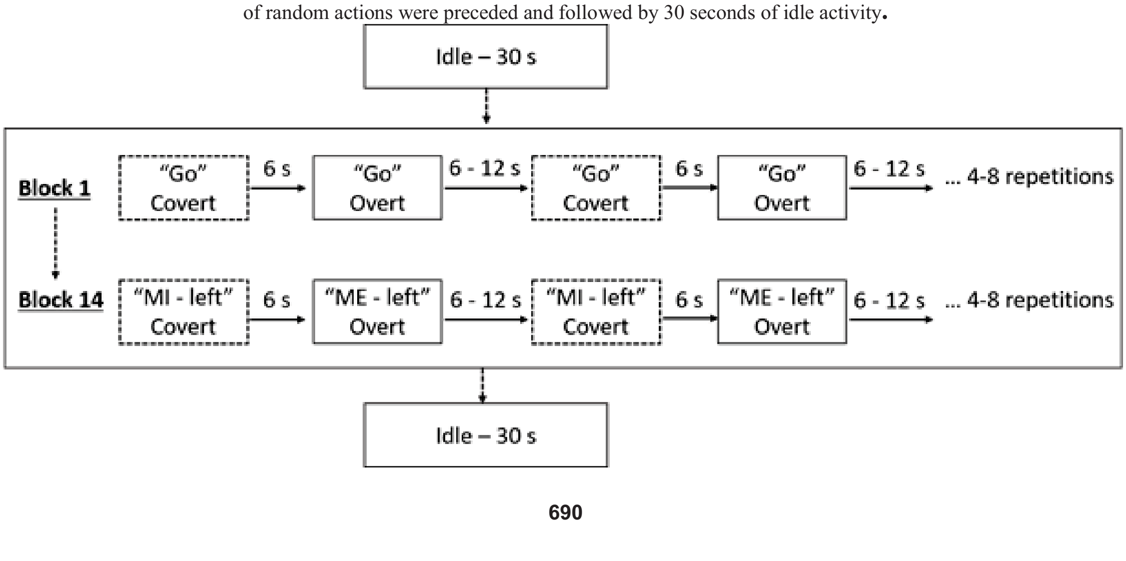


D. Data analysis

All data processing was performed in MATLAB 2017a.

Pre-processing: Before feature extraction, all channels were visually inspected. Channels influenced by large artefacts were excluded for the respective subject. For each subject an average of 2.0 ± 2.2 channels was excluded from the analysis. For all features, the EEG was divided into epochs containing data from three seconds before (to capture the Bereitschaftspotential [20]), to two seconds after action onset (to capture the action). EMG was filtered with a passband of 20-200 Hz and a dynamic threshold was used to identify the ME and overt speech onsets; starting at 80 % of the maximal

Figure 1. Timeline. Each action was performed covertly and repeated overtly after 6 seconds. The next covert repetition was self-paced but occurred in the range 6-12 seconds. An action was repeated covertly and overtly in blocks of 4-8 repetitions, and a total of 14 blocks of random actions were preceded and followed by 30 seconds of idle activity.



EMG amplitude in the signal and incrementally decreasing until all onsets were found (80 of each action).

Feature extraction: Temporal features were simple means of the EEG when filtered with a passband of 0.5-4 Hz in 500 milliseconds segments without overlap. Spectral features were the spectral power (root-mean-square) within six frequency bands, in one-second segments without overlap. The six frequency bands were: delta (0.5 - 4 Hz), theta (4 - 8 Hz), alpha (8 - 13 Hz), beta (13 - 30 Hz), gamma (30 - 45 Hz) and high gamma (55 - 80 Hz). The temporal and spectral features resulted in a total of 40 features per channel. All the features were derived separately for all 25 channels, thus a maximum of 1000 features (range: 760-1000) could be derived, unless channels were rejected from further analysis.

Classification: After feature extraction, the epochs of each class were randomly divided into training and test sets in a 10-fold cross-validation. All classification accuracies reported are the average of the 10 test sets. The following classification problems were tested: speech (4 classes), movement (3 classes) and combined speech and movement (6 classes). The features were classified using a random forest classifier [21]. The random forest was implemented with 128 trees.

III. RESULTS

The results are summarized in the confusion matrices (**Figure 3-8**). All values are the mean ± standard deviation across subjects in percent. In all the three classification scenarios the class of idle activity has the highest classification accuracy.

A. Speech

The average classification accuracy for covert speech was 67 ± 9 % (**Figure 3**) which was lower than overt speech with classification accuracy of 75 ± 7 % (**Figure 4**).

Figure 3. Confusion matrix for the covert speech classes and the idle class

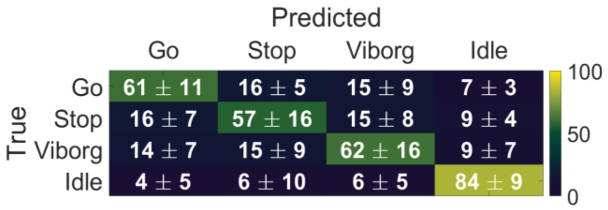
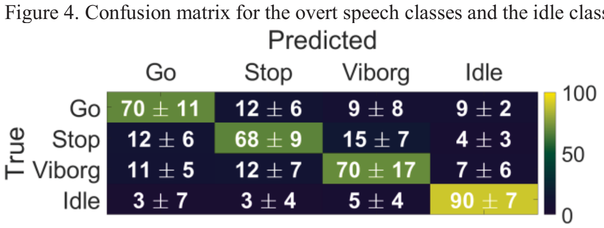


Figure 4. Confusion matrix for the overt speech classes and the idle class.



B. Movement

The average classification accuracy for MI was 77 ± 6 % (**Figure 5**), which was similar to the classification accuracy for movement execution of 79 ± 6 % (**Figure 6**).

Figure 5. Confusion matrix for the movement imagery classes and the idle class.

Figure 6. Confusion matrix for the movement execution classes and the idle class.

C. Combination of speech and motor classes

The average classification accuracy for all covert actions (speech and MI) was 61 ± 9 % (**Figure 7**), which was lower than the average classification accuracy for all overt actions (speech and movement execution) of 67 ± 7 % (**Figure 8**).

Figure 7. Confusion matrix for the covert speech classes, the movement imagery classes, and the idle class.

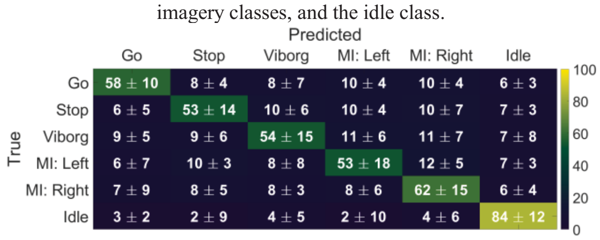
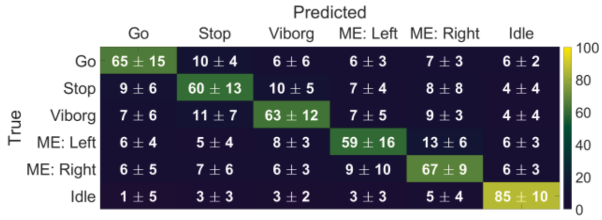


Figure 8. Confusion matrix for the overt speech classes, the movement execution classes, and the idle class.



IV. DISCUSSION

The aim of the present study was to investigate the feasibility of classifying covert speech in addition to motor-imagery actions for controlling an external device. By using spectral power and simple mean as features and a random forest classifier, both MI and covert speech could be classified with a better than random [22] average classification accuracy in all three classification scenarios. The performance of covert speech and MI in the current study is comparable to, but slightly lower than the performance obtained with overt speech and motor execution which was expected [23]. The highest classification accuracies were obtained for the motor classes, but it was also expected that the classification accuracy would decrease when more classes were added. The only class that differed from other classes was idle. This indicates that speech classes were as feasible as motor classes in this BCI-system and usable in a combined speech-motor controlled BCI. Moreover, this is also supported by the good classification accuracy in the combined scenario with six classes.

The overall classification accuracies reached in this study are comparable to or better than previous studies using covert speech as control signal. In a two-class paradigm, Nguyen et al. (2018) tested different BCI systems on covert speech [24]. In that study, an average classification accuracy of 80.1 % was achieved in the best case [24]. In an ECoG study with covert and overt speech of vowels and consonants, a classification accuracy of 37.5 % and 36.9 % was reached, for covert speech of vowels and consonants respectively [18]. Studies investigating BCIs using different types of mental tasks as control signals have reached a classification accuracy of 62 % with a 4-class EEG BCI [13], and 78 % in a 3-class BCI [14]. However, it should be noted that there are methodological differences. Despite the high classification accuracy, it may not be enough to operate a wheelchair or robotic arm satisfactorily [25]. The features were extracted from 5-second epochs which greatly limits the amount of control commands that can be sent in a given time window to the external device [26]. It was not tested if the epoch length could be reduced, but this should be tested in future studies to be able to send more commands to the external devices where fast control may be needed such as in wheelchair control.

In the current study, only seven healthy subjects participated, and the proposed framework is thereby still in its infancy to draw conclusions about the feasibility of using speech classes in a BCI system for people with spinal cord injuries or similar motor deficits. In future studies the usability of the speech approach should be tested. In addition, the BCI system results are based on offline classification with synchronized action onsets, most likely resulting in a higher classification accuracy than what could have been reached in an online system. Due to the self-paced nature of this study, synchronization of epoch onset might have been off in some cases, possibly having a negative influence on accuracy.

V. CONCLUSION

The classification accuracy obtained for covert speech was above chance level which indicates that this approach could be used for BCI control and a potential alternative for patients who cannot use MI. Moreover, it was possible to combine speech and motor classes and maintain a high classification accuracy. It is necessary to perform a usability test to investigate if covert or overt speech can be used as a more intuitive BCI control paradigm alone, or in conjunction with motor classes.

REFERENCES

[1] J. L. Collinger, M. L. Boninger, T. M. Bruns, K. Curley, W. Wang, and D. J. Weber, “Functional priorities, assistive technology, and brain-computer interfaces after spinal cord injury,” *J. Rehabil. Res. Dev.*, vol. 50, no. 2, pp. 145–160, 2013.

[2] J. R. Wolpaw and C. B. Boulay, “Brain Signals for Brain–Computer Interfaces,” in *Brain–Computer Interfaces*, B. Graimann, B. Allison, and G. Pfurtscheller, Eds. Springer, 2010, pp. 29–46.

[3] L. Bi, X. Fan, and Y. Liu, “EEG-Based Brain-Controlled Mobile Robots : A Survey,” *IEEE Trans. Human-Machine Syst.*, vol. 43, no. 2, pp. 161–176, 2013.

[4] J. Minguillon, M. A. Lopez-Gordo, and F. Pelayo, “Trends in EEG-BCI for daily-life: Requirements for artifact removal,” *Biomed. Signal Process. Control*, vol. 31, pp. 407–418, 2017.

[5] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, “Brain Computer Interfaces for communication and control,” *Clin. Neurophysiol.*, vol. 113, pp. 767–791, 2002.

[6] A. Kübler *et al.*, “The thought translation device: A neurophysiological approach to communication in total motor paralysis,” *Exp. Brain Res.*, vol. 124, no. 2, pp. 223–232, 1999.

[7] C. W. Chen, M. S. Ju, Y. N. Sun, and C. C. K. Lin, “Model analyses of visual biofeedback training for EEG-based brain-computer interface,” *J. Comput. Neurosci.*, vol. 27, no. 3, pp. 357–368, 2009.

[8] E. V Friedrich, R. Scherer, and C. Neuper, “Long-term evaluation of a 4-class imagery-based brain-computer interface,” *Clin. Neurophysiol.*, vol. 124, no. 5 PG-916-927, pp. 916–927, 2013.

[9] G. Pfurtscheller, C. Guger, G. Müller, G. Krausz, and C. Neuper, “Brain oscillations control hand orthosis in a tetraplegic,” *Neurosci. Lett.*, vol. 292, no. 3, pp. 211–214, 2000.

[10] G. Muller-Putz *et al.*, “Towards Noninvasive Hybrid Brain–Computer Interfaces: Framework, Practice, Clinical Application, and Beyond,” *Proc. IEEE*, vol. 103, no. 6, pp. 926–943, 2015.

[11] B. Z. Allison, J. Jin, Y. Zhang, and X. Wang, “A four-choice hybrid P300/SSVEP BCI for improved accuracy,” *Brain-Computer Interfaces*, vol. 1, no. 1, pp. 17–26, 2014.

[12] C. Sannelli, C. V. Id, K. Mu, and B. Blankertz, *A large scale screening study with a SMR- based BCI : Categorization of BCI users and differences in their SMR activity*, no. Mi. 2019.

[13] E. V Friedrich, C. Neuper, and R. Scherer, “Whatever works: a systematic user-centered training protocol to optimize brain-computer interfacing individually,” *PLoS One*, vol. 8, no. 9, p. e76214, 2013.

[14] R. Chai, S. H. Ling, G. P. Hunter, Y. Tran, and H. T. Nguyen, “Classification of Wheelchair Commands using Brain Computer Interface: Comparison between Able-Bodied Persons and Patients with Tetraplegia,” *35th Annu. Int. Conf. IEEE EMBS*, 2013.

[15] X. Pei, J. Hill, and G. Schalk, “Silent communication: Toward using brain signals,” *IEEE Pulse*, vol. 3, no. 1, pp. 43–46, 2012.

[16] S. Martin, I. Iturrate, J. del R. Millán, R. T. Knight, and B. N. Pasley, “Decoding inner speech using electrocorticography: Progress and challenges toward a speech prosthesis,” *Front. Neurosci.*, vol. 12, no. JUN, pp. 1–10, 2018.

[17] F. H. Guenther and J. S. Brumberg, “Brain-machine interfaces for real-time speech synthesis.,” *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE Eng. Med. Biol. Soc.*, vol. 2011, no. PG-5360-3, pp. 5360–5363, 2011.

[18] X. Pei, D. D. Barbour, E. C. E. Leuthardt, and G. Schalk, “Decoding vowels and consonants in spoken and imagined words using electrocorticographic signals in humans.,” *J. Neural Eng.*, vol. 8, no. 4, p. 46028, 2011.

[19] F. Pulvermüller, W. Lutzenberger, and H. Preissl, “Nouns and verbs in the intact brain: evidence from event-related potentials

and high-frequency cortical responses.," *Cereb. Cortex*, vol. 9, pp. 497–506, 1999.

[20] J. J. McArdle, Z. Mari, R. H. Pursley, G. M. Schulz, and A. R. Braun, "Electrophysiological evidence of functional integration between the language and motor systems in the brain: a study of the speech Bereitschaftspotential," *Clin. Neurophysiol.*, vol. 120, pp. 275–284, 2009.

[21] L. Breiman, "Random Forests," pp. 1–33, 2001.

[22] G. R. Müller-putz, R. Scherer, C. Brunner, R. Leeb, and G. Pfurtscheller, "Better than random ? A closer look on BCI results," *Int. J. Bioelectromagn.*, vol. 10, no. 1, pp. 52–55, 2008.

[23] M. Jochumsen, I. K. Niazi, N. Mrachacz-Kersting, N. Jiang, D. Farina, and K. Dremstrup, "Comparison of spatial filters and features for the detection and classification of movement-related cortical potentials in healthy individuals and stroke patients," *J. Neural Eng.*, vol. 12, no. 5, p. 10pp, 2015.

[24] C. H. Nguyen, G. Karavas, and P. Artemiadis, "Inferring imagined speech using EEG signals: a new approach using Riemannian Manifold features," *J. Neural Eng.*, vol. 15, p. 16pp, 2018.

[25] O. Tonet *et al.*, "Defining brain-machine interface applications by matching interface performance with device requirements," *J. Neurosci. Methods*, vol. 167, no. 1 PG-91-104, pp. 91–104, 2008.

[26] P. Yuan, X. Gao, B. Allison, Y. Wang, G. Bin, and S. Gao, "A study of the existing problems of estimating the information transfer rate in online brain-computer interfaces," *J. Neural Eng.*, vol. 10, no. 2, 2013.