Decoding Semantic Categories from EEG Activity in Silent Speech Imagination Tasks

1st Maurice Rekrut
German Research Center for Artificial Intelligence (DFKI)
Saarbrücken, Germany
Maurice.Rekrut@dfki.de

2nd Mansi Sharma
German Research Center for Artificial Intelligence (DFKI)
Saarbrücken, Germany
Mansi.Sharma@dfki.de

3rd Matthias Schmitt
German Research Center for Artificial Intelligence (DFKI)
Saarbrücken, Germany
Matthias.Schmitt@dfki.de

4th Jan Alexandersson
German Research Center for Artificial Intelligence (DFKI)
Saarbrücken, Germany
Jan.Alexandersson@dfki.de

5th Antonio Krüger
German Research Center for Artificial Intelligence (DFKI)
Saarbrücken, Germany
Antonio.Krueger@dfki.de

***Abstract*—Silent Speech Brain-Computer Interfaces try to decode imagined or silently spoken speech from brain activity. This technology holds big potential in various application domains, e.g. restoring communication abilities for handicapped people, or in settings where overtly spoken speech is not an option due to environmental conditions, e.g. noisy industrial or aerospace settings. However, one major drawback of this technology still is the limited number of words which can be distinguished at a time. This work therefore introduces the concept of Semantic Silent Speech BCIs, which adds a layer for semantic category classification prior to the actual word classification to multiply the number of classifiable words in Silent Speech BCIs many times over. We evaluated the possibilities of classifying 5 different semantic categories of words during a word imagination task by comparing various feature extraction and classification methods. Our results show remarkable classification accuracies of up to 95% for the single best subject with a Common Spatial Pattern (CSP) feature extraction and a Support Vector Machine (SVM) classifier and a best average classification accuracy of 60.44% for a combination of CSP and a Random Forrest (RF) classifier. Even a cross-subject analysis over the data of all subjects lead to results above the chance level of 20%, with a best performance of 43.54% for a self assembled feature vector and a RF classifier. Those results clearly indicate that the classification of the semantic category of an imagined word from EEG activity is possible and therefor lay the foundation for Semantic Silent Speech BCIs in the future.**

***Index Terms*—Silent Speech, BCI, EEG, Semantic Processing**

I. Introduction

Silent Speech Interfaces (SSI) are defined in Human-Computer Interaction as the the concept of speech communication in the absence of an audible acoustic signal [1] and have become a widely researched topic in the field of Brain-Computer Interfaces (BCIs) over the last years [2]–[4]. Studies showed that it is possible to decode imagined words from brain activity measured invasively at the surface of the brain [5], [6] but even with non-invasive measures like Electroencephalography (EEG) [7]–[10]. One major drawback of the existing non-invasive approaches for such an alternative communication pathway however, is the maximum number of distinguishable words. The presented approach in [11] achieved a classification accuracy of 70% for a three word classification problem on EEG data, which makes those approaches appear applicable even in real world scenarios. However, three words result in limited possibilities concerning communication. As soon as the number of words increases the classification accuracy decreases significantly. [12] reported a classification accuracy of 58.41% for 5 silently spoken words and [13] even managed to classify 12 words from EEG activity with an accuracy of around 34.2%. These results are astonishing and clearly above chance level, but far below the expectations towards a classifier applied for a real world communication and application.

We propose to bridge this gap arising from the trade-off between number of classifiable words and classification accuracy in EEG-based SS-BCIs by a 2-layer approach similar to those mentioned in [14], [15] and start with a classification of the semantic category of a word before proceeding to classifying the word itself. Some researchers have been exploring this direction in the past: [14] used a limited set of 2 categories and the results of [15] are questionable due to their study setup as pointed out in [16] but the concept itself has the potential to multiply the number of classifiable words in Silent Speech BCIs many times over.

In one of our last studies we showed the feasibility of classifying words from 5 different semantic categories, namely action words, living, locational, non-living, and numbers, in object based decision tasks [17]. The participants had to answer

978-1-7281-4707-9/20/$31.00 ©2020 IEEE

simple yes/no-questions concerning the words in order to trigger conscious processing. The results yield a classification accuracy of 84.61% for the single best subject for a Common Spatial Pattern (CSP) feature extraction method and Random Forrest (RF) classifier and 57,29% classification accuracy on average with the same setup. Those results clearly indicated the potential of this method for the use in SS-BCIs. In this study we went one step further and classified semantic categories during actual word imagination in order to prove the feasibility in SS-BCI applications.

With this paper we provide 3 primary contributions to the field of Silent Speech BCIs:

- We introduce the concept of Semantic Silent Speech BCIs which makes use of semantic classification prior to word classification in order to increase the number of detectable words in SS-BCIs.
- We show the feasibility of classifying semantic categories from EEG activity for the first time for 5 different semantic categories during actual word imagination.
- We provide recommendations for the best setup for semantic category detection in SS-BCIs by comparing the performance of different feature extraction and classification methods.

II. Material and Methods

The goal of this study was to classify semantic categories of imagined words from EEG activity and to compare the results to our previous study for classification of semantic categories in object based decision tasks. [17] This study therefore followed a similar setup and used the same data analysis methods in order to make the results of both studies comparable. The following section gives an overview over the material and methods used and introduces the concept of the Semantic Silent Speech BCI.

A. Semantic Silent Speech BCI

The concept of the Semantic Silent Speech BCI tries to increase the number of distinguishable words of a Silent Speech BCI by integrating a second layer of semantic category classification prior to the actual word classification layer. The idea is based on a study of Huth et al [18] which showed, that it is possible to cluster patterns of similar brain activity for words with the same semantic category while the participants were listening to short stories with those words embedded. Those similar patterns could not only be shown for the individual but also among the brain activity of all subjects which indicates that it might be possible to train a classifier on certain patterns of brain activity for semantic categories and achieve single-trail classification. The main advantage and purpose in the scope of our work however lies on the increased number of words which might be classified by adding this additional semantic layer to Silent Speech BCIs.

Let n be the number of different semantic categories and m be the number of different words equally distributed among those categories in a Silent Speech task. With the additional semantic layer, we can divide the original m-class classification task (resulting from the overall m numbers of words) into a n-class classification task followed by an $\frac{m}{n}$-class classification task. Given the fact that the semantic processing in the brain was shown to be present spread over the whole cortex [18] and that imagined speech production mainly evokes the left hemisphere around Wernicke's area [19], the feature space and sources of those two related cognitive processes, speech production and semantic processing, should be separable and provide distinguishable features. By decomposing the original n-class classification task into smaller classification tasks with a separated feature space, we expect the number of classifiable words to increase by n (= number of semantic categories) in the best case.

B. Subjects

The study was conducted with 20 healthy subjects (age 21–29). All subjects were native German speakers who were right-handed and had a normal or corrected-to-normal vision. Subjects where chosen to have the same mother tongue, to prevent confounding neurolinguistic effects on the EEG due to foreign language use [20], and prevent multilingual requirements to the setup and subject population [21]. The subjects were asked not to consume caffeinated substances at least three hours before the starting of data collection as they have a proven potential to affect the EEG recordings [22]. Each subject was introduced to the task, and informed consent was obtained from all subjects for scientific use of the recorded data. The data was acquired in a dim light room with minimized distractions like external sound, mobile devices and others, where the voluntary participants were asked to sit in a comfortable chair to prevent unnecessary muscle movements to reduce noise and artifacts in the EEG, which could emerge from mental stress, unrelated sensory input, physiological motor activity and electrical interference. Three subjects (1,2 and 19) were excluded due to poor electrode-to-skin contact later, leaving a total of 17 subjects (8 male, 9 female).

C. Recording

EEG signals were recorded using a wireless 32 channel electroencephalograph system namely g.Nautilus with g.Scarabeo electrodes (g.tec medical engineering GmbH, Austria). The sampling rate was set to 500 Hz. The 10-20 International System of electrode placement was used to locate the electrodes. This configuration is believed to cover the whole scalp resulting in the capturing of spatial information from the brain recordings effectively which provided the optimal setup for our study based on the findings of [18].

D. Study Setup

The objective of this study was to decode semantic categories of imagined words using EEG signals. Therefor the task included the imagination of words from different semantic categories. The overall experimental paradigm for recording the silent speech part is depicted in Figure 1. Each trial starts with a fixation cross shown in the time interval of t_1. Subsequently, the word was presented on the computer screen,

which was later followed by the blinking of the fixation cross once. Then the participant was instructed to imagine the word by making use of sub-vocalization, i.e., saying the word in your mind once. A short break of 2s followed the next trial.
The same 5 semantic categories were selected as during our last study [17], originally based on the findings of [18], namely: living, non-living, numbers, locations and action verbs. Each category contained 10 words to be processed, Table I gives an detailed overview.
In order to prevent classifying arbitrary brain states based on block-level temporal correlations rather than stimulus-related activity [23], we chose the partial-block-wise presentation paradigm which was introduced in [17]. In this procedure the words were presented in blocks of 10 according to their category. The words inside the blocks were randomly shuffled, each block was presented randomly once per trail and the experiment consisted of 4 trails, resulting in a total number of 200 samples per participant.

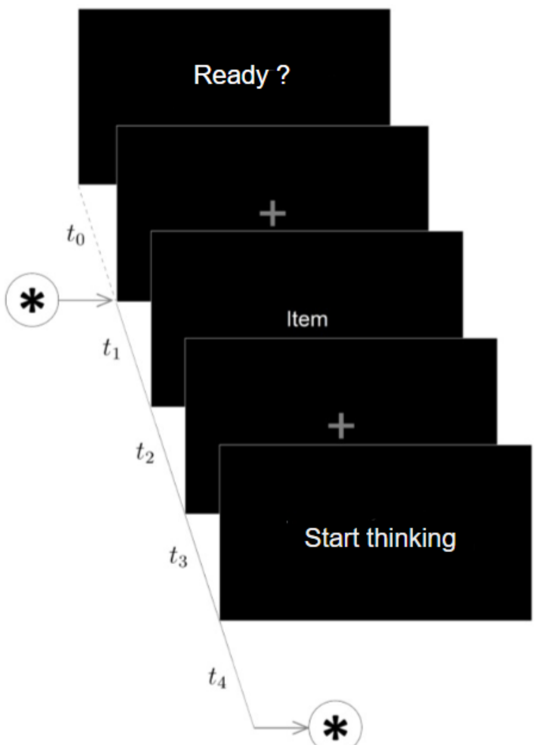

Fig. 1. Procedure of visual output on the screen during the silent speech imagination task, as done for 5 semantic categories in each trial ($t_0 = \infty$, $t_1 = (15. \pm 5.)s$, $t_2 = 2s$, $t_3 = .30s$ and $t_4 = 2.5s$).

E. Data analysis

Preprocessing A basic preprocessing, including filtering and referencing, was applied to remove unwanted artifacts. Channels containing obvious signal quality issues as well as data segments identified to contain strong artifacts were labeled bad and thus, excluded from further analysis. The EEG was filtered with a second-order Butterworth notch filter with a lower cutoff frequency of 48Hz and an upper cutoff frequency of 52Hz, to remove power line noise. A fourth-order Butterworth

TABLE I
SELECTED SEMANTIC CATEGORIES AND ITEMS PRESENTED IN THE SILENT SPEECH TASK, ORIGINALLY IN GERMAN, TRANSLATED TO ENGLISH.

Category	Items
Locational	Kitchen, Bathroom, Cellar, Garden, Court, Bedroom, Living room, Staircase, Corridor, Attic
Actions	Throw, Open, Lift, Lower, Switch on, Switch off, Put, Place, Push, Pull
Living	Dog, Cat, Peacock, Lamb, Pigeon, Mother, Father, Grandmother, Grandfather, Doctor
Non-living	Light, Shutters, Heater, Television, Telephone, Computer, Stove, Refrigerator, Washing machine, dryer
Numbers	One, Two, Three, Four, Five, Six, Seven, Eight, Nine, Ten, Eleven

was applied for high pass filtering at a cutoff frequency of 1Hz as recommended by [24]. Lowpass filtering was done using a 17th order Chebyshev type 2 filter with a cutoff frequency of 200Hz preserving the frequency range up to high-gamma [25], and a minimum of 60dB attenuation in the stop band, to obtain adequate roll-off. All the electrodes were referenced to common average over all electrodes to achieve low signal-to-noise ratio [26]. Further denoising of the EEG was done using a joined method of Independent Component Analysis (ICA) and wavelet denoising, referred to as wavelet-enhanced independent component analysis (wICA) [27] to remove artifacts.
Epoching The continuous EEG signal was cut into smaller epochs of different length for further analysis. Epochs were extracted in reference to the stimulus onset from the pre-processed data. To further analyze the impact of different epoching duration, three intervals were chosen, namely, $T_1 = 0.3s - 0.8s$, $T_2 = 0.3s - 1.5s$ and $T_3 = 0s - 2s$. T_1 and T_2 were supposed to cover the point around 400 ms after stimulus onset, as studies on semantics assume this to be the earliest point information is consciously processed [14]. The third interval was chosen as the entire period between stimulus onset and end of the task in order to cover as much information as possible.
Feature Extraction We investigated two feature extraction methods, first we assembled a feature vector containing features from the time and frequency domain and based on the different state-of-the-art literature in silent speech detection [28]–[31]. The features were extracted with the open-source python module PyEEG [32]. As a second feature extraction method we chose Common Spatial Patterns (CSP) which is a frequently used technique in BCI applications [33], [34].
Originally the CSP algorithm is developed for binary classification problems, but there are some studies showing the multiclass classification with One-vs-all scheme [35]–[37]. In this study, we implemented the multiclass classification setup as mentioned in [17] with five labels and a chance level at 20%.

Classification Classification was done based on two strategies, cross-subject and within-subject. In the cross-subject analysis, the data from all subjects is taken together as a single input in order to find similar patterns among subjects while in within-subject analysis, the performance of the classifier is computed on the individual subjects data set. We used Support Vector Machine (SVM) and Random Forest (RF), because of their frequent use in EEG-based Silent Speech BCIs [7], [10], [38], both evaluated using grid search and five-fold cross-validation with a test train split of 0.1. For the cross-subject data set, we further used a deep artificial neural network, the Multilayer Perceptron using the ADAM and SGD optimizer with mean-squared error as loss function. In this case grid search was performed to find the optimal number of hidden layers and chose the best learning rate in addition to the activation function. Unfortunately, within-subject analysis with neural networks was not possible due to the limited availability of data of an individual subject. All the experiments were multi-class classification problems with five labels resulting in a chance level of 20%.

Performance metrics To evaluate the performance of the classifiers, we have calculated the classification accuracy. Accuracy is the most commonly used performance metric and defined as the ratio of the total number of correct predictions to the total number of predictions overall. We have furthermore created the confusion matrix and classification reports including F1-score, Precision and Recall and present them in this work for the best performing subjects.

III. Results

Table II shows the results for the cross-subject data set evaluated with the different classifiers (SVM, RF, MLP), feature extraction methods (CSP, FV) and concerning the different epoching intervals (T_1, T_2, T_3). For the CSP feature extraction condition, the classifiers did not manage to exceed chance level. In all other cases, the classifiers evened out around 40% which is significantly above chance level and RF classifier achieved the highest accuracy for all epoching intervals.

TABLE II
MEAN ACCURACY FOR THE CROSS-SUBJECT CLASSIFICATION DEPENDING ON THE DIFFERENT EPOCHING INTERVALS, FEATURE EXTRACTION METHODS (CSP AND FV) AND CLASSIFIERS (SVM, RF, MLP) USED.

Classifier	T_1 = 0.3s - 0.8s	T_2 = 0.3s - 1.5s	T_3 = 0s - 2s
SVM-CSP	$19.93 \pm 1.90\%$	$20.05 \pm 2.11\%$	$20.94 \pm 1.35\%$
RF-CSP	$22.54 \pm 2.64\%$	$20.64 \pm 2.07\%$	$21.65 \pm 1.24\%$
SVM-FV	$38.45 \pm 2.27\%$	$37.21 \pm 2.06\%$	$41.59 \pm 1.79\%$
RF-FV	$42.18 \pm 1.61\%$	$40.05 \pm 2.25\%$	$43.54 \pm 2.13\%$
MLP-ADAM	$38.92 \pm 2.75\%$	$38.69 \pm 2.77\%$	$39.64 \pm 1.10\%$
MLP-SGD	$40.17 \pm 1.08\%$	$40.94 \pm 3.74\%$	$40.88 \pm 2.50\%$

Figures 2 – 5 show the results for the within-subject condition and the different classifiers used on the assembled feature vector and the CSP feature extraction methods. For within-subject analysis using the support vector machine and the feature vector, Figure 2 summarizes the results for all three intervals. Overall, the highest accuracy of 68.42% is achieved by subject 4 in the duration T_1. The mean accuracies were reported to be 46.08%, 45.81%, and 47.81% in the duration of T_1, T_2 and T_3 respectively. Except for subjects 5 and 6, all the other subjects managed to achieve classification results above chance level as represented by the blue horizontal line across all the data arrangements and epoching intervals.

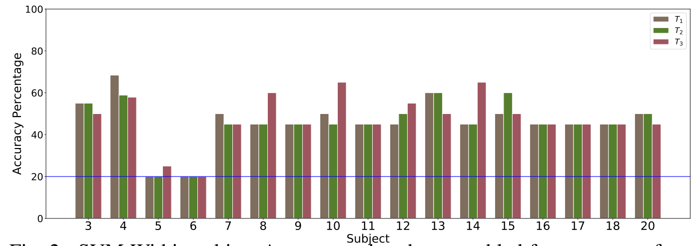

Fig. 2. SVM Within-subject Accuracy using the assembled feature vector for three different epoching intervals, $T_1 = 0.3s - 0.8s$, $T_2 = 0.3s - 1.5s$, $T_3 = 0s - 2s$

Figure 3 shows the Accuracy versus Subject plot for all three epoching intervals using the random forest classifier and the feature vector. The overall highest accuracy of 70% is achieved by subjects 16 and 11 in the epoching interval T_3 and T_2, respectively. The mean accuracies are 52.81%, 52.52%, and 55.18% in the duration of T_1, T_2 and T_3 respectively. In this case, al the subjects managed to score above chance level, i.e., 20% represented by the horizontal blue line. As compared to SVM, RF reported better accuracies for all the subjects. Figure 4 presents the results obtained with the implemented

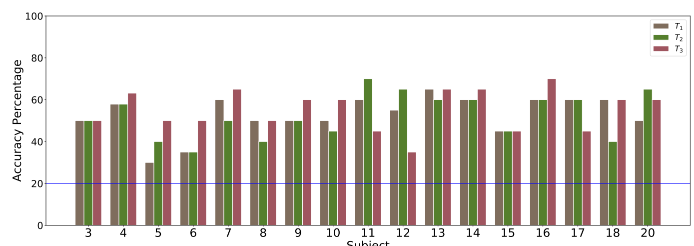

Fig. 3. RF Within-subject Accuracy using the assembled feature vector for three different epoching intervals, $T_1 = 0.3s - 0.8s$, $T_2 = 0.3s - 1.5s$, $T_3 = 0s - 2s$

CSP algorithm using a support vector machine. The overall highest accuracy of **95%** is achieved by subject 10 in the interval of T_3. The confusion matrix for this condition can be found in figure 6 and the classification report in table III. The highest accuracies of 90%, 85%, and **95%** are achieved in the interval T_1, T_2, and T_3, respectively. The mean accuracies are 56.91%, 51.08%, and 55.77% in interval of T_1, T_2 and T_3 respectively. The worst performing subject is subject 13, in the interval T_1. However, in the other two intervals T_2 and T_3, subject 13 manages to achieve accuracies above chance level.

Figure 5 illustrates within-subject accuracy for all three epoching duration's using random forest with horizontal line representing the chance level. Overall, the highest accuracy achieved is 90% by subject 20 in the epoching interval of $T_1 = 0.3s - 0.8s$. The confusion matrix for this condition can be found in figure 7 and the classification report in table

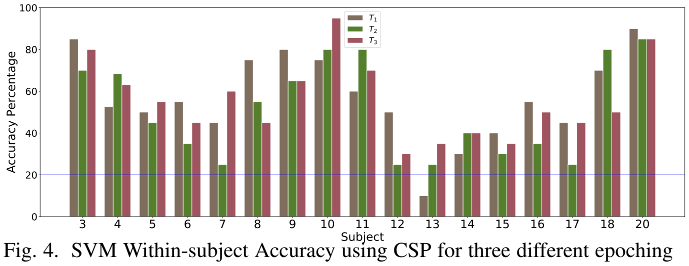

Fig. 4. SVM Within-subject Accuracy using CSP for three different epoching intervals, $T_1 = 0.3s - 0.8s$, $T_2 = 0.3s - 1.5s$, $T_3 = 0s - 2s$

TABLE III
CLASSIFICATION REPORT FOR SUBJECT 10 USING SVM-CSP AND T_3

Labels	F1-Score	Precision	Recall	Support
actions	0.85	0.75	1	3
living	1	1	1	6
Non-living	1	1	1	4
Locational	0.67	1	0.5	2
Numbers	1	1	1	5

IV. The mean accuracies are 60.44%, 51.43%, and 51.93%, in T_1, T_2, and T_3, respectively. The worst performing subject is 13, below chance level in the interval T_2. However, in other two intervals T_1 and T_3, subject 13 manages to achieve above chance level. Subject 15 is at chance level in the interval T_2, while in the interval T_1, subject 15 manages to reach an accuracy of 60%. Therefore, we can observe a lot of fluctuation in accuracy in different epoching durations.

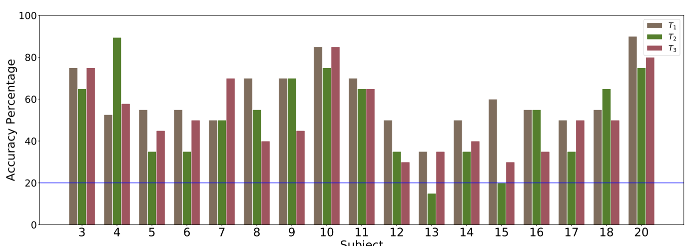

Fig. 5. RF Within-subject Accuracy using CSP for three different epoching intervals, $T_1 = 0.3s - 0.8s$, $T_2 = 0.3s - 1.5s$, $T_3 = 0s - 2s$

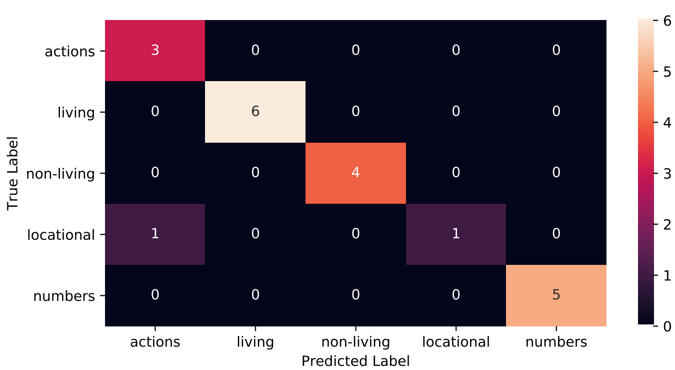

Fig. 6. Confusion Matrix for Subject 10 in the interval T_3 using SVM-CSP

IV. DISCUSSION

The main goal of this study was to classify EEG activity related to semantic processing during word imagination. In

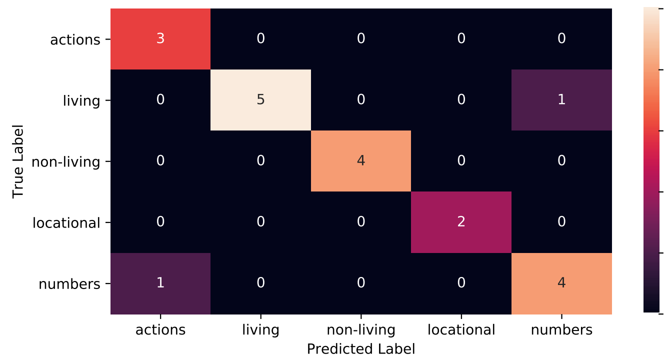

Fig. 7. Confusion Matrix for Subject 20 in the interval T_1 using RF-CSP

TABLE IV
CLASSIFICATION REPORT FOR SUBJECT 20 USING RF-CSP AND T_1

Labels	F1-Score	Precision	Recall	Support
actions	0.85	0.75	1	3
living	0.90	1	0.83	6
Non-living	1	1	1	4
Locational	1	1	1	2
Numbers	0.8	0.8	0.8	5

our attempt to train different classifier and feature extraction methods on the recorded EEG data of all participants in the cross-subject condition, we can say, that the assembled feature vector as well as the Multi-Layer-Perceptron approach managed to exceed chance level witch classification accuracies of up to 43.54% for the Random Forrest Classifier and the assembled Feature Vector. These results indicate that the classification of semantic categories of imagined words based on cross-subject data might be possible. This hypothesis should be further investigated however with a larger dataset, especially to make use of the potential of the neural networks. The hypothesis, that the imagination of words from the same semantic category produces similar spatial patterns of EEG activity across all subjects, as shown for fMRI measurements in [18], could not be proven by the Common Spatial Pattern method, with classification accuracy around the chance level of 20%. These results are consistent with the findings of our previous study for the classification of semantic categories during object based decision tasks. [17]

The within-subject condition yield similar results as well. Our analysis clearly indicates that it is possible to classify the semantic category of a word a person was subvocalizing in his/her head with a remarkable single best classification accuracy of 95% for subject 10 with CSP feature extraction and SVM classifier. However, there does not seem to be one best setup or combination of epoching, feature extraction or classification methods. Comparing the two feature extraction methods used we can say that CSP deliverd better results on average than the assembled feature vector. While the feature vector managed to produce a more uniform distribution amongst the subjects with no significant outliers, the results for CSP vary strongly among subjects.

The two classifiers used do not produce significantly different results within the two feature extraction methods but again on average the Random Forrest classifier performed slightly better than the Support Vector Machine with one exception for SVM and CSP at T_3 where the single best subject was reported.

Although those results show that on average the best method for semantic category classification in word imagination tasks (among those presented in this study) is a Common Spatial Pattern feature extraction combined with a Random Forrest Classifier with an epoching interval of 300 - 800 ms after word imagination onset, we clearly recommend to tailor those methods to the individual. As shown in our previous study for object based decision tasks, the word imagination task appears to be highly subject specific. While subject 14 managed to achieve classification accuracies of above 60% for the Feature Vector it hardly manages to exceed chance level for the Common Spatial Pattern method. The opposite holds for subjects 5 and 6 although improvement can be shown when switching from a Support Vector Machine to a Random Forrest Classifier in the Feature Vector condition. While subject 10 achieved the best results in all feature extraction and classification methods within the epoching interval T_3, subject 3 performed best within the interval T_1, and the results of subject 4 regarding the epoching interval vary completely under the different conditions.

V. Conclusion and future work

The goal of this work was to explore the potential of classifying the semantic categories of imagined words from EEG data for the use in Silent Speech BCIs. We introduced the concept of Semantic Silent Speech BCIs and implemented various feature extraction and classification methods as well as different epoching intervals with the aim to provide a recommendation for a best possible setup concerning semantic category classification during word imagination. We furthermore analysed the data in a cross-subject approach with the intention to find similar patterns in brain activity among subjects but also investigated a within-subject condition where the classification was done on the data of each subject individually.

The cross-subject results did not show the expected common spatial patterns among all participants but promising results could be achieved on temporal and spectral features. With classification accuracies of up to 43.54% the proposed method with a Random Forrest classifier clearly exceeded the chance level of 20% but is still far from an accuracy needed for real world applications. This approach should be further evaluated in the future on a larger data set to fully exploit the potential of Neural Networks for classification which usually work with a tremendous amount of input data, larger than what could be provided in this study.

The results for the individual subjects were promising as well but highly distributed among the different epoching intervals, feature extraction and classification methods. The results clearly indicate that it is possible to classify the semantic category of a word during word imagination, with a best average classification accuracy of 60.44% for CSP feature extraction and SVM classifier in a time interval of 0.3s - 0.8s after imagination onset, and a best single subject classification accuracy of even 95% for CSP and RF classifier for the full length time interval of 0s - 2s. As shown in our last study on semantic category classification in object based decision tasks, the epoching intervals, feature extraction and classification methods are highly subject specific. There was no clear best setup to recommend for semantic category classification but rather the conclusion to select the different methods tailored to the individual based on predefined training sessions.

As future work we plan to extend our collected datasets and go the last step on the way to a real Semantic Silent Speech BCI by actually implementing the 2-layer approach of classifying the semantic category of an imagined word first followed by the classification of the word itself in an applied Semantic Silent Speech BCI scenario.

VI. Acknowledgment

References

[1] João Freitas, António Teixeira, Miguel Sales Dias, and Samuel Silva. *An Introduction to Silent Speech Interfaces*. Springer, 2017.

[2] Bruce Denby, Tanja Schultz, Kiyoshi Honda, Thomas Hueber, Jim M Gilbert, and Jonathan S Brumberg. Silent speech interfaces. *Speech Communication*, 52(4):270–287, 2010.

[3] Stephanie Martin, Peter Brunner, Iñaki Iturrate, José del R Millán, Gerwin Schalk, Robert T Knight, and Brian N Pasley. Word pair classification during imagined speech using direct brain recordings. *Scientific reports*, 6:25803, 2016.

[4] Mashael M AlSaleh, Mahnaz Arvaneh, Heidi Christensen, and Roger K Moore. Brain-computer interface technology for speech recognition: A review. In *2016 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA)*, pages 1–5. IEEE, 2016.

[5] Stephanie Martin, Peter Brunner, Iñaki Iturrate, José del R Millán, Gerwin Schalk, Robert T Knight, and Brian N Pasley. Word pair classification during imagined speech using direct brain recordings. *Scientific reports*, 6:25803, 2016.

[6] Christian Herff, Dominic Heger, Adriana De Pesters, Dominic Telaar, Peter Brunner, Gerwin Schalk, and Tanja Schultz. Brain-to-text: decoding spoken phrases from phone representations in the brain. *Frontiers in neuroscience*, 9:217, 2015.

[7] Li Wang, Xiong Zhang, Xuefei Zhong, and Yu Zhang. Analysis and classification of speech imagery eeg for bci. *Biomedical signal processing and control*, 8(6):901–908, 2013.

[8] Chad M Spooner, Erik Viirre, and Bradley Chase. From explicit to implicit speech recognition. In *International Conference on Augmented Cognition*, pages 502–511. Springer, 2013.

[9] Torres-García Alejandro Antonio, Reyes-García Carlos Alberto, and Villaseñor-Pineda Luis. Toward a silent speech interface based on unspoken speech. In *Proceeding of the 5th International Joint Conference on Biomedical Engineering Systems and Technologies*, 2012.

[10] Mohanchandra K and Saha S. A communication paradigm using subvocalized speech: translating brain signals into speech. *Augmented Human Res 1 3*, 2016.

[11] Chuong H Nguyen, George K Karavas, and Panagiotis Artemiadis. Inferring imagined speech using eeg signals: a new approach using riemannian manifold features. *Journal of neural engineering*, 15(1):016002, 2017.

[12] Erick F González-Castañeda, Alejandro A Torres-García, Carlos A Reyes-García, and Luis Villaseñor-Pineda. Sonification and textification: Proposing methods for classifying unspoken words from eeg signals. *Biomedical Signal Processing and Control*, 37:82–91, 2017.

[13] Seo-Hyun Lee, Minji Lee, Ji-Hoon Jeong, and Seong-Whan Lee. Towards an eeg-based intuitive bci communication system using imagined speech and visual imagery. In *2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)*, pages 4409–4414. IEEE, 2019.

[14] Taekyung Kim, Jeyeon Lee, Hoseok Choi, Hojong Lee, In-Young Kim, and Dong Pyo Jang. Meaning based covert speech classification for brain-computer interface based on electroencephalography. In *2013 6th International IEEE/EMBS Conference on Neural Engineering (NER)*, pages 53–56. IEEE, 2013.

[15] Pradeep Kumar, Rajkumar Saini, Partha Pratim Roy, Pawan Kumar Sahu, and Debi Prosad Dogra. Envisioned speech recognition using eeg sensors. *Personal and Ubiquitous Computing*, 22(1):185–199, 2018.

[16] Ren Li, Jared S Johansen, Hamad Ahmed, Thomas V Ilyevsky, Ronnie B Wilbur, Hari M Bharadwaj, and Jeffrey Mark Siskind. Training on the test set? an analysis of spampinato et al.[arxiv: 1609.00344]. *arXiv preprint arXiv:1812.07697*, 2018.

[17] Maurice Rekrut, Mansi Sharma, Matthias Schmitt, Jan Alexandersson, and Antonio Krüger. Decoding semantic categories from eeg activity in object-based decision tasks. In *2020 8th International Winter Conference on Brain-Computer Interface (BCI)*, pages 1–7. IEEE, 2020.

[18] Alexander G Huth, Wendy A De Heer, Thomas L Griffiths, Frédéric E Theunissen, and Jack L Gallant. Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature*, 532(7600):453, 2016.

[19] Akihiko Tsukahara, Masayuki Yamada, Keita Tanaka, and Yoshinori Uchikawa. Analysis of eeg frequency components and an examination of electrodes localization during speech imagery. In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 4698–4702. IEEE, 2019.

[20] Roland H Grabner, Clemens Brunner, Robert Leeb, Christa Neuper, and Gert Pfurtscheller. Event-related eeg theta and alpha band oscillatory responses during language translation. *Brain research bulletin*, 72(1):57–65, 2007.

[21] Barbara Höhle. *Psycholinguistik*. Walter de Gruyter, 2011.

[22] J. Lass M. Saifudinova, M. Bachmann and H. Hinrikus. Effect of coffee on eeg spectral assymmetry. *In World Congress on Medical Physics and Biomedical Engineering*, 2015.

[23] Ren Li, Jared S. Johansen, Hamad Ahmed, Thomas V. Ilyevsky, Ronnie B. Wilbur, Hari M. Bharadwaj, and Jeffrey Mark Siskind. Training on the test set? an analysis of spampinato et al. [31]. *CoRR*, abs/1812.07697, 2018.

[24] Irene Winkler, Stefan Debener, Klaus-Robert Müller, and Michael Tangermann. On the influence of high-pass filtering on ica-based artifact reduction in eeg-erp. In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 4101–4105. IEEE, 2015.

[25] Nathan E Crone, Alon Sinai, and Anna Korzeniewska. High-frequency gamma oscillations and human brain mapping with electrocorticography. *Progress in brain research*, 159:275–295, 2006.

[26] Kip A Ludwig, Rachel M Miriani, Nicholas B Langhals, Michael D Joseph, David J Anderson, and Daryl R Kipke. Using a common average reference to improve cortical neuron recordings from microelectrode arrays. *Journal of neurophysiology*, 101(3):1679–1689, 2009.

[27] Laurent Albera, Amar Kachenoura, Pierre Comon, Ahmad Karfoul, Fabrice Wendling, Lotfi Senhadji, and Isabelle Merlet. Ica-based eeg denoising: a comparative analysis of fifteen methods. *Bulletin of the Polish Academy of Sciences: Technical Sciences*, 60(3):407–418, 2012.

[28] Anaum Riaz, Sana Akhtar, Shanza Iftikhar, Amir Ali Khan, and Ahmad Salman. Inter comparison of classification techniques for vowel speech imagery using eeg sensors. In *The 2014 2nd International Conference on Systems and Informatics (ICSAI 2014)*, pages 712–717. IEEE, 2014.

[29] Gautam Krishna, Yan Han, Co Tran, Mason Carnahan, and Ahmed H Tewfik. State-of-the-art speech recognition using eeg and towards decoding of speech spectrum from eeg. *arXiv preprint arXiv:1908.05743*, 2019.

[30] Shunan Zhao and Frank Rudzicz. Classifying phonological categories in imagined and articulated speech. In *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 992–996. IEEE, 2015.

[31] Beomjun Min, Jongin Kim, Hyeong-jun Park, and Boreom Lee. Vowel imagery decoding toward silent speech bci using extreme learning machine with electroencephalogram. *BioMed research international*, 2016, 2016.

[32] Forrest Sheng Bao, Xin Liu, and Christina Zhang. Pyeeg: an open source python module for eeg/meg feature extraction. *Computational intelligence and neuroscience*, 2011, 2011.

[33] Javeria Khan, Muhammad Hamza Bhatti, Usman Ghani Khan, and Razi Iqbal. Multiclass eeg motor-imagery classification with sub-band common spatial patterns. *EURASIP Journal on Wireless Communications and Networking*, 2019(1):174, 2019.

[34] Thanh Nguyen, Imali Hettiarachchi, Amin Khatami, Lee Gordon-Brown, Chee Peng Lim, and Saeid Nahavandi. Classification of multi-class bci data by common spatial pattern and fuzzy system. *IEEE Access*, 6:27873–27884, 2018.

[35] J. Muller-Gerking H. Ramoser and G. Pfurtscheller. Optimal spatial filtering of single trial eeg during imagined hand movement. *IEEE Trans. Rehab. Eng.*, vol. 8, pp. 441–446, 2000.

[36] Soumyadip Chatterjee, Saugat Bhattacharyya, Amit Konar, D. N. Tibarewala, Anwesha Khasnobish, and Ramadoss Janarthanan. Performance analysis of multiclass common spatial patterns in brain-computer interface. In *Pattern Recognition and Machine Intelligence - 5th International Conference, PReMI 2013, Kolkata, India, December 10-14, 2013. Proceedings*, pages 115–120, 2013.

[37] Moritz Grosse-Wentrup and Martin Buss. Multiclass common spatial patterns and information theoretic feature extraction. *IEEE Trans. Biomed. Engineering*, 55(8):1991–2000, 2008.

[38] Juan Manuel Mayor Torres, Evgeny A Stepanov, and Giuseppe Riccardi. Eeg semantic decoding using deep neural networks. In *Rovereto Workshop on Concepts, Actions, and Objects (CAOS)*, 2016.