Proceedings Paper:
AlSaleh, M., Moore, R., Christensen, H. et al. (2018) Discriminating between imagined speech and non-speech tasks using EEG. In: 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) , 18-21 Jul 2018, Honolulu, Hawaii. IEEE, pp. 1952-1955. ISBN: 978-1-5386-3646-6. ISSN: 1558-4615.

https://doi.org/10.1109/EMBC.2018.8512681

Discriminating between Imagined Speech and Non-Speech Tasks using EEG

Mashael AlSaleh,[1] Roger Moore[1], Heidi Christensen[1], Mahnaz Arvaneh[2]

Abstract— **People who are severely disabled (e.g Locked-in patients) need a communication tool translating their thoughts using their brain signals. This technology should be intuitive and easy to use. To this line, this study investigates the possibility of discriminating between imagined speech and two types of non-speech tasks related to either a visual stimulus or relaxation. In comparison to previous studies, this work examines a variety of different words with only single imagination in each trial. Moreover, EEG data are recorded from a small number of electrodes using a low-cost portable EEG device. Thus, our experiment is closer to what we want to achieve in the future as communication tool for locked-in patients. However, this design makes the EEG classification more challenging due to a higher level of noise and variations in EEG signals. Spectral and temporal features, with and without common spatial filtering, were used for classifying every imagined word (and for a group of words) against the non-speech tasks. The results show the potential for discriminating between each imagined word and non-speech tasks. Importantly, the results are different between subjects using different features showing the need for having subject specific features.**

I. Introduction

A Brain Computer Interface (BCI) is a communication system that directly translates brain signals to control commands without requiring any muscular activities. BCI can be potentially the only communication option for people who suffer from severe neuromuscular impairments such as locked-in patients. Many instructed cognitive tasks have been explored for BCI ranging from selective attention, motor imagery, words associations, to mental arithmetic [1]. The use of these modalities for communication can be limiting as they are unintuitive [1], limited in the number of classes that can be provided (e.g. only four classes from motor imagination studies [2], and/or requiring external stimuli (e.g. P300-based BCIs)). In order to have a communication technology that is more close to reading thoughts, a considerable literature has grown up recently around the theme of detecting imagined speech. Imagined speech can be defined as asking subjects to imagine the pronunciation of words as if they were pronouncing it aloud, but without any articulator movements. On contrast to other instructed cognitive tasks such as motor imagery, detecting speech imagination is still a new research domain and a lot of questions were not completely answered and identified. This includes: the optimal experimental design, important brain areas to capture brain activities related to speech, and the effect of phonological and semantic differences between words in the recognition.

In majority of BCI studies, electroencephalogram (EEG) was used as a non-invasive tool to record brain activities. EEG is portable, relatively affordable, and has a good temporal resolution. However, EEG signals include superfluous noise, redundant unwanted information and a poor spacial resolution. Importantly, these weaknesses are even more pronounced when using non-gel wireless EEG devices. Studies in the area of imagined speech using EEG technologies, can be divided into three types, based on the type of imagined speech used, namely word imagination [3], [4], [5], [6], [7], syllable imagination [8], [9] and vowel imagination [10], [11].

Few studies have focused on discriminating between imagined speech and non-speech. One study included a comparison between the imagination of two vowels (/a/,/u/) by imagine lips movement and 'no imagination' as a control state [10], [12]. In Zhao *et al.* [9], the authors investigated three mental states related to speech imagination, actual speech and stimulus presentation (a word presented on the screen and a sound utterance played). This study facial expressions and audio signals were combined with EEG signals for improving classification results. In [7], [13], the authors used EEG recorded from 10 seconds word repetitions of yes and no versus unconstrained rest time.

On contrary to literature, this paper targets a more intuitive imagined speech procedure. This includes imagining words once rather than several times in a fixed time window. Moreover, the imagination involves a larger variety of words (i.e. 11 words and syllables). Finally, a low cost wireless EEG head set is used for recording brain signals. All these factors imply larger variations in imagined speech EEG signals making it more difficult to classify. To this line, this paper focuses on classification between the imagined words versus either relaxation or the attention to a visual stimulus. Spatio-spectral and time domain features are examined for each subject to extract the information form EEG signals. We explore different time intervals for feature extraction. The results are presented as first how words as a group can be classified from the non-speech class using the proposed features and classification algorithms. Then, the potential of classifying each individual word versus relaxation is discussed.

* This research has been supported by the Saudi Ministry of Education, King Saud University, Saudi Arabia, and University of Sheffield, UK.

[1]M. Alsaleh, R. Moore, H. Christensen are with the Dept. of Computer Science, University of Sheffield, Uk. emails: (mmalsaleh1, r.k.moore, heidi.christensen)@sheffield.ac.uk

[2] M. Arvaneh is with the Dept.of Automatic Control and Systems Engineering at the University of Sheffield, email: m.arvaneh@sheffield.ac.uk

II. Experiment

A. Participants

Nine males ranging in age from 18 to 36 participated in this study. Participants with any neurological disorders, a history of brain injury, a personal or family history of epilepsy or those who had consumed alcohol or any types of drug in the previous 12 hours were excluded from the study.

B. EEG data acquisition

The acquisition of brain signals was done by using the Emotiv EEG neuro-headset. This headset has a total of 16 channels; 14 channels were used for data recording (i.e. AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, AF4) and two were inactivated as the ground and reference channels.

C. Stimuli

In this study, eleven words were selected based on variations in their semantic meaning. Several neuroscience studies have examined the impact of the emotional implications of words on neural activities as represented by cortical potentials [14], [15]. Syllables were chosen for the 'no semantic' stimuli, which is the approach used in previous studies [8]. In the present study, the word stimuli were selected to include emotional words, words with neutral meaning (directions and responses) and syllables, as follows:

- Syllables: "/ba/" and "/ka/".
- Directions:"Left", "Right", "Up", "Down".
- Responses: "Yes", "No".
- Emotions: "Happy", "Sad", "Help".

D. Task

Before starting to record the EEG signals, the experimental instructions were explained to each participant. These instructions were written out as a script to ensure consistency between all nine participants. The instructions asked the participants to minimise their body movements during the experiment. It was explained that this included hand movements, jaw movements and any other kind of physical movement. The task steps and the stimuli presented are summarised below:

1) **Visual attention**(fixation): The symbol, '+', was presented on a screen for one second. The participants were instructed to focus on the symbol.
2) **Relaxation** (black screen): In this task, the participants were instructed to relax (be silent) and clear their minds from any type of thinking as much as possible. This task lasted two seconds.
3) **Word presentation:** In this task, a word was presented on the screen for two seconds. The presentation of words from this list was done randomly to avoid the effect of word order.
4) **Word imagination** (black screen): Once the screen gets blank, the participants were instructed to immediately imagine the previously presented word for one time. This task lasted two seconds.

A total of 11 imagined speech stimuli were used. The recording was done as blocks. Six blocks were recorded for each subject. During each block, each word was presented in random order eight times. Hence, a total of 88 fixations and relaxation tasks were conducted for each block (they were presented before and after each word). A total of 48 trials were recorded for each word; and all the stimuli in the experiment consisted of 1584 trials.

III. Data Analysis

A. Data pre-processing

High-pass and low-pass zero-phase filters were applied in the range of 1–30 Hz to remove power line noise, and attenuate noise caused by body movements. For all nine subjects, the F7 and F8 channels were used as the ground channels and the AF4 and AF3 channels were removed because they are near the eyes, and most signals recorded from them were related to eye blinking and movement [16]. Moreover, baseline correction was done to remove the effects that occurred prior to the presentation of each stimulus. The baseline can be defined as the time preceding the stimulus. Here, we removed -200 ms to 0 ms with respect to the stimulus onset [17].

B. Feature extraction

In the feature extraction stage, we investigated the spatio-spectral and temporal features. Time domain features were extracted by computing four features from each channel: Standard Deviation (SD), Mean, Sum of Values (SUM), Root mean Square (RMS). Spatial features were computed using the Filter-bank Common Spatial Patterns (CSP) algorithm. Both spatio-spectral and temporal features were calculated for three different time intervals after the start of the task: [0-1s], [0-1.5 s], and [0- 2 s].

1) Time domain features: The proposed time domain features have been used in the literature in several EEG studies. For example, in [18], SD, RMS, SUM, and Energy have been used to classify envisioned speech.

In this study SD, RMS, SUM, and mean are calculated for the samples resulting in 4 features from each channel. As we used 12 channels, it led to 48 time domain features.

2) Spatio-spectural features: EEG data have poor spatial resolution; therefore, in order to discriminate between the two classes it was necessary to design some spatial filters. Common spatial patterns (CSP) is a well-known spatial filtering algorithm that are based on maximising the variance of one class while minimising it for the other class [19].

Let a single trial EEG be represented as $E \epsilon R^{c \times s}$, where c denotes the channels and s samples. The CSP algorithm filters the matrix E to X, given as:

$$X = EW \tag{1}$$

The spatial filter W is a projection matrix that was computed based on simultaneous diagonalization of the covariance matrices from both classes [19]. As in [19], not all the spatial filtered signals were used for extracting features. Instead, only a defined number, m, of the first and last rows of X in

(1) are used for feature extraction. In the present study, m is equal to 2. Assuming the signals X_p $(p = 1.....2m)$ are given, the feature vector F is calculated as:

$$F_p = log(var(X_p)/\sum_{i=1}^{2m} var(X_p)) \quad (2)$$

However, CSP may lead to poor classification accuracies if the data is inappropriately filtered with the wrong frequency bands. In [20], Ang *et al.* proposed that applying a filter bank that filters EEG data into multiple bands can improve the results. Seven filters were included in the bank to obtain data ranging between 1 Hz and 30 Hz. This frequency range represents the well-known bands in the literature, and it has been interpreted as delta, theta, alpha, low beta, mid-beta, high beta and low gamma. Since the EEG data was filtered using seven frequency bands, and four rows of the CSP filtered signals were considered for each band, the total number of spatio-spectral features was 28.

C. Classification

The two groups of proposed features (i.e. time domain and spatio-spectral) were evaluated separately in different trial length: [0- 1s], [0- 1.5s], [0- 2s]. For both groups, using train data, Pearson correlations between features and class labels were calculated to rank features. For the classification, 8-fold cross validation was applied to divide the data into training (80%), (10%) development data and (10%) testing data. The development set was used to identify the best number of features for every subjects. The linear discriminant analysis (LDA) algorithm and Linear Support Vector Machine (SVM) were used as classification algorithms.

IV. Results and discussion

A. Speech vs non-speech for group of words

EEG trials related to word imagination were labeled as one group and classified against the non-speech tasks (either visual attention or relaxation). The number of speech trials is 528 trials and the same number is for each of the non-speech tasks. The visual attention is related to two stimuli, '+' and word presentation.

As shown in Table I and II, on average the classification accuracies between visual attention and imagined speech was better than the classification accuracies between the imagined speech and relaxation across all classifiers in all time intervals when filter bank CSP features were used. This makes sense as visual attention provokes visual processing in brain that is absent in speech imagery and relaxation. The maximum classification accuracy for visual attention is in the time [0-1 s]. Importantly, in the classification between speech and relax state, all subjects except S3, S4, and S9 achieved classification accuracies above %60 using filter-bank CSP.

B. Spatio-spectural features vs Time domain features to classify imagined words vs relaxation

In Table III, we computed time domain features to classify between imagined words versus relaxation. The average accuracy across subjects in both classifiers and all time intervals was less than the classification of filter-bank CSP features. However, the results was different between subjects. For example, for S1, S2 and S3 time domain features yielded less accurate results, whereas for S4 time domain features yielded better results. This suggest that applying feature selection method to select between time-domain and spatio-spectral features would further enhance the results.

TABLE I

Average classification accuracy (%) between relaxing (non-speech) and all imagined words using Filter-bank CSP features

	SVM			LDA		
Subject	[1s]	[1.5s]	[2s]	[1s]	[1.5s]	[2s]
S1	69	70	70	70	70	72
S2	68	68	62	68	66	62
S3	57	58	58	55	57	58
S4	57	58	58	55	57	58
S5	61	62	60	61	62	59
S6	66	65	66	67	65	67
S7	62	59	61	62	59	59
S8	63	61	59	63	60	60
S9	59	57	59	58	57	58
Average	62.1	62	61.3	62.1	62	61.3
SD	4.72	4.53	3.94	5.17	4.22	4.71

TABLE II

Average classification accuracy (%) between visual attention (non-speech) and all imagined words using Filter-bank CSP features

	SVM			LDA		
Subject	[1s]	[1.5s]	[2s]	[1s]	[1.5s]	[2s]
S1	67	67	65	67	66	65
S2	75	77	73	75	78	72
S3	71	67	65	71	67	65
S4	58	56	55	59	58	55
S5	73	66	59	72	62	58
S6	69	65	65	69	65	64
S7	62	61	59	60	60	60
S8	68	70	67	65	70	67
S9	60	56	55	58	56	54
Average	67	65	62.6	66.2	64.7	62.2
SD	5.9	6.7	6	6.1	6.7	5.9

TABLE III

Average classification accuracy (%) between relaxing (non-speech) and all imagined words using time domain features

	SVM			LDA		
Subject	[1s]	[1.5s]	[2s]	[1s]	[1.5s]	[2s]
S1	48	49	49	56	54	53
S2	63	66	64	70	67	65
S3	52	48	53	52	48	53
S4	59	57	57	63	65	66
S5	59	57	57	63	65	66
S6	61	63	61	63	62	63
S7	51	52	53	57	55	56
S8	67	66	67	67	65	67
S9	51	53	50	53	54	50
Average	56.3	56.3	56.4	59.4	58.4	58.7
SD	6.5	7	6.3	6.5	6.5	6.5

C. Classification of individual words versus relaxation

Each individual word was imagined in 48 trials during the experiment. In the classification of each word versus relaxation, classification between the 48 trials of the imagined word were compared with the 48 trials of relaxation that happened before the same word. Table IV and Table V show the classification accuracies using CSP and time domain features using SVM and LDA and different trials lengths. The results are very encouraging as we used only a small number of training trials, a low-cost EEG device, and single imagination repetition. In comparison with the results in Table I and III, interestingly the classification of all words as one group can help in identifying the best classifier and type of best features for each subject. For example, for S1 form Table I we can find that the best time interval is [0-1 s] which is consistent with Table IV. For S4, S2 we can conclude that time domain features using LDA can give best number of classified words.

TABLE IV

NUMBER OF WORDS THAT PROVIDE ABOVE CHANCE LEVEL CLASSIFICATION ACCURACY (%60) AGAINST RELAXATION USING FILTER-BANK CSP FEATURES

	SVM			LDA		
Subject	[1s]	[1.5s]	[2s]	[1s]	[1.5s]	[2s]
S1	4	4	11	2	7	7
S2	4	3	3	5	4	3
S3	0	1	2	0	0	2
S4	0	1	0	1	3	0
S5	2	1	1	5	1	1
S6	3	3	2	5	3	3
S7	3	2	3	4	1	2
S8	3	4	4	4	4	3
S9	0	1	2	1	1	2

TABLE V

NUMBER OF WORDS THAT PROVIDE ABOVE CHANCE LEVEL CLASSIFICATION ACCURACY (%60) AGAINST RELAXATION USING TIME DOMAIN FEATURES

	SVM			LDA		
Subject	[1s]	[1.5s]	[2s]	[1s]	[1.5s]	[2s]
S1	0	1	1	3	2	5
S2	3	4	4	7	7	6
S3	0	1	2	1	1	2
S4	2	0	1	4	5	4
S5	0	1	1	1	2	1
S6	8	4	2	5	8	5
S7	2	1	3	4	3	3
S8	9	9	10	11	9	9
S9	1	0	0	3	2	2

V. CONCLUSION AND SUMMARY

This study is a first step in understanding how imagined speech can be recognised from another tasks using only EEG data and single imagination of the word. The study examined listed of varied stimuli. We have shown that the results vary across subjects and according to different types of tasks. Moreover, the contribution of features is different depending on the task. In this study we have not find any differences between stimuli in terms of classification accuracy. In future work, we will examine different types of features and their combinations to improve the results.

REFERENCES

[1] M. Van Gerven, J. Farquhar, R. Schaefer, R. Vlek, J. Geuze, A. Nijholt, N. Ramsey, P. Haselager, L. Vuurpijl, S. Gielen *et al.*, "The brain–computer interface cycle," *Journal of neural engineering*, vol. 6, no. 4, p. 041001, 2009.

[2] C. Brunner, R. Leeb, G. Müller-Putz, A. Schlögl, and G. Pfurtscheller, "Bci competition 2008–graz data set a," *Institute for Knowledge Discovery (Laboratory of Brain-Computer Interfaces), Graz University of Technology*, vol. 16, 2008.

[3] M. Wester, "Unspoken speech-speech recognition based on electroencephalography," *Master's Thesis, Universitat Karlsruhe (TH)*, 2006.

[4] A. Porbadnigk, M. Wester, and T. S. Jan-p Calliess, "Eeg-based speech recognition impact of temporal effects," 2009.

[5] L. Wang, X. Zhang, X. Zhong, and Y. Zhang, "Analysis and classification of speech imagery eeg for bci," *Biomedical signal processing and control*, vol. 8, no. 6, pp. 901–908, 2013.

[6] E. F. González-Castañeda, A. A. Torres-García, C. A. Reyes-García, and L. Villaseñor-Pineda, "Sonification and textification: Proposing methods for classifying unspoken words from eeg signals," *Biomedical Signal Processing and Control*, vol. 37, pp. 82–91, 2017.

[7] A. R. Sereshkeh, R. Trott, A. Bricout, and T. Chau, "Eeg classification of covert speech using regularized neural networks," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 25, no. 12, pp. 2292–2300, 2017.

[8] M. DZmura, S. Deng, T. Lappas, S. Thorpe, and R. Srinivasan, "Toward eeg sensing of imagined speech," in *International Conference on Human-Computer Interaction*. Springer, 2009, pp. 40–48.

[9] S. Zhao and F. Rudzicz, "Classifying phonological categories in imagined and articulated speech," in *Acoustics, Speech and Signal Processing (ICASSP), 2015 IEEE International Conference on*. IEEE, 2015, pp. 992–996.

[10] C. S. DaSalla, H. Kambara, M. Sato, and Y. Koike, "Single-trial classification of vowel speech imagery using common spatial patterns," *Neural networks*, vol. 22, no. 9, pp. 1334–1339, 2009.

[11] L. Sarmiento, P. Lorenzana, C. Cortes, W. Arcos, J. Bacca, and A. Tovar, "Brain computer interface (bci) with eeg signals for automatic vowel recognition based on articulation mode," in *Biosignals and Biorobotics Conference (2014): Biosignals and Robotics for Better and Safer Living (BRC), 5th ISSNIP-IEEE*. IEEE, 2014, pp. 1–4.

[12] N. Yoshimura, A. Satsuma, C. S. DaSalla, T. Hanakawa, M.-a. Sato, and Y. Koike, "Usability of eeg cortical currents in classification of vowel speech imagery," pp. 1–2, 2011.

[13] A. R. Sereshkeh, R. Trott, A. Bricout, and T. Chau, "Online eeg classification of covert speech for brain–computer interfacing," *International journal of neural systems*, vol. 27, no. 08, 2017.

[14] C. Herbert, M. Junghofer, and J. Kissler, "Event related potentials to emotional adjectives during reading," *Psychophysiology*, vol. 45, no. 3, pp. 487–498, 2008.

[15] A. Schacht and W. Sommer, "Time course and task dependence of emotion effects in word processing," *Cognitive, Affective, & Behavioral Neuroscience*, vol. 9, no. 1, pp. 28–43, 2009.

[16] S. S. Gupta, S. Soman, P. G. Raj, R. Prakash, S. Sailaja, and R. Borgohain, "Detecting eye movements in eeg for controlling devices," in *Computational Intelligence and Cybernetics (CyberneticsCom), 2012 IEEE International Conference on*. IEEE, 2012, pp. 69–73.

[17] G. F. Woodman, "A brief introduction to the use of event-related potentials in studies of perception and attention," *Attention, Perception, & Psychophysics*, vol. 72, no. 8, pp. 2031–2046, 2010.

[18] P. Kumar, R. Saini, P. P. Roy, P. K. Sahu, and D. P. Dogra, "Envisioned speech recognition using eeg sensors," *Personal and Ubiquitous Computing*, pp. 1–15, 2017.

[19] H. Ramoser, J. Muller-Gerking, and G. Pfurtscheller, "Optimal spatial filtering of single trial eeg during imagined hand movement," *IEEE transactions on rehabilitation engineering*, vol. 8, no. 4, pp. 441–446, 2000.

[20] K. K. Ang, Z. Y. Chin, H. Zhang, and C. Guan, "Filter bank common spatial pattern (fbcsp) in brain-computer interface," pp. 2390–2397, 2008.