Postprint of: García-Salinas J. S., Villaseñor-Pineda L., Reyes-García C. A., Torres-García A. A., Transfer learning in imagined speech EEG-based BCIs, Biomedical Signal Processing and Control, Vol. 50 (2019), pp. 151-157, DOI: 10.1016/j.bspc.2019.01.006

Transfer Learning in Imagined Speech EEG-based BCIs

Jesús S. García-Salinas[a], Luis Villaseñor-Pineda[a], Carlos A. Reyes-García[a], Alejandro Torres-García[a]

[a]*Biosignals Processing and Medical Computing Laboratory,*
Instituto Nacional de Astrofísica Óptica y Electrónica,
Luis Enrique Erro No. 1, Santa María Tonatzintla, Puebla, México

Abstract

The Brain-Computer Interfaces (BCI) based on electroencephalograms (EEG) are systems which aim is to provide a communication channel to any person with a computer, initially it was proposed to aid people with disabilities, but actually wider applications have been proposed. These devices allow to send messages or to control devices using the brain signals. There are different neuro-paradigms which evoke brain signals of interest for such purposes. Imagined speech is one of the most recent paradigms, and it is explored in this work, it consists of the internal pronunciation of a word, i. e., a subject imagines the utterance of a word without emitting sounds or articulating facial movements. Under this neuro-paradigm, to increase the initial vocabulary reducing drastically the training time using few or none new data is an open challenge. The proposed method extracts characteristic units (i. e. *codewords*) of the EEGs associated with the words of an initial vocabulary. Subsequently, a new imagined word is represented with these codewords, and then a classification algorithm is applied. The method was tested both, with and without calibration examples, in a 27 subjects dataset. An initial vocabulary of 4 words, with 33 epochs for each word was considered. The results were obtained by averaging the accuracies of every subject, without calibration data a 65.65% accuracy was achieved. In comparison to the baseline method, which obtained an average accuracy of 68.9%, the proposed method showed no statistical difference.

Keywords: Imagined Speech, Bag of features, EEG, Brain Computer Interfaces, Transfer learning
2010 MSC: 00-01, 99-00

1. Introduction

A Brain-Computer Interface (BCI) is a system which can be used to transform the brain signals into commands to control a device. For this task, the user must produce a brain activity pattern, which can be evoked internally or produced by an external stimulus. In this work electroencephalograms (EEG) were used to record brain electrophysiological activity.

There are different neuro-paradigms based on internal evocations of brain activity. Imagined movement or motor imagery is a widely used BCI neuro-paradigm in different types of BCIs. However, it has a limitation due to the availability of imagined movements (i.e. the number of extremities to move). In this work, another paradigm based on an internal stimulus is used: imagined speech. Imagined speech consists of the internal pronunciation of a word or phoneme, i. e. a subject imagines the utterance of a word without emitting sounds or articulating facial movements.

Under this neuro-paradigm, the first challenge is to recognize an initial vocabulary. Previous works have addressed this task [1, 2, 3, 4] following different speech approaches, v. gr. phonemes, syllables or words. So far, only small vocabularies (a maximum of 6 words) have been analyzed. Nevertheless, an advantage of this paradigm is precisely the possibility of using large vocabularies (of dozens of words), which consider the expressiveness of the language. However, by using a traditional supervised learning scheme, it is necessary to have training data for each word to be learned. The acquisition of this data is a slow and very demanding process for the users. Hence, to provide methods which add new words without increasing the cost in the training phase is an area of interest.

The present work takes the first steps in this direction. The main idea is to verify whether is possible to represent a new word using the elements calculated for previous words (i.e. an initial vocabulary). For this, a Bag of Features (BoF) approach is applied. This method consists in calculate a *codebook* which contains a set of *codewords* that represents the EEG signals of the initial vocabulary [5, 6, 7]. Later on, these codewords will represent a new imagined word. With this representation a classifier is trained, it will be able to discriminate the original vocabulary and the new word. In this way, the cost of the training phase will decrease by reducing the number of examples needed to recognize the new word.

2. Related works

2.1. Imagined speech

Actually, most of the BCIs are developed to recognize the brain activity related to motor imagery tasks. A different approach to BCI control is the imagined speech task. Previous works have evidenced that imagined speech can be discriminated in EEG signals [8, 1, 2, 9]. The main advantage of this approach is the use of the language expressiveness.

Due to the experimental designs, acquisition protocols and imagined speech vocabularies; previous works differ not only on the proposed method but on the evaluation. In [10] a set of words was labeled according to its phonemic and phonological features, the experimental set includes EEG, face and audio recordings. Many schemes of binary classification were explored like vowel vs consonant, presence of nasal, presence of bilabial, presence of high-front vowel and presence of high back vowel, the best recognition rate using only EEG recordings was 63.5% and obtained under the presence of nasals with an SVM-quad classifier. This data includes sixty-four channels from twelve subjects.

Recently, in [8] different wavelet families and classifiers were explored for multi-class imagined speech classification, the dataset was conformed by five Spanish words (”Arriba”, ”Abajo”, ”Izquierda”, ”Derecha”, ”Seleccionar”) corresponding to (”Up”, ”Down”, ”Left”, ”Right”, ”Select”) and twenty seven subjects. Also an automatic channel selection based on fuzzy inference were implemented to reduce the dataset, and an accuracy of $68.18 \pm 16\%$ were achieved. In [1], a larger vocabulary was presented, it includes the Spanish words (”Arriba”, ”Abajo”, ”Izquierda”, ”Derecha”, ”Adelante”, ”Atras”) translated as (”Up”, ”Down”, ”Left”, ”Right”, ”Forward”, ”Backward”) which were recorded from fifteen subjects, and applied the method presented on [11], the obtained accuracy rate was $18.58 \pm 1.47\%$. The low accuracy was related to the randomization of the stimuli in the acquisition protocol and to the use of basic spectral features extraction. Nevertheless, the presence of imagined speech discriminative data in the EEG signals were not discarded.

2.2. Bag of Features

This method is based on the *Vector Quantization* traditional approach, which objective is to achieve an automatic signal characterization discretizing its representation. In the signal analysis area, many variations and adaptations of the method have been developed and it receives different names according to the application area, is referred to as a bag of words in document classification, bag of instances in multiple instance learning, bag of frames in audio and speech recognition, bag of patterns in signal processing and pattern recognition and bag of images or visual words in computer vision [12].

Bag of Features (BoF) has been applied in widespread areas of application and has been improved and modified to fulfill different requirements.

In [13], the bag of patterns representation was applied to classify electrocardiogram (EKG) time series using hierarchical and partitional clustering, to create a bag of patterns representations. It may be noticed that the feature extraction, pre-processing and classification steps, are independent steps of the bag of patterns.

In the specific case of EEG signals, the bag of features was used by [14], where EEG and EKG signals were analyzed for epilepsy detection. Features were extracted from a one channel EEG applying a DWT, these features were clustered by the k-means algorithm. Then, a set of histograms were created and finally classified by the 1-Nearest Neighbor algorithm. The accuracy obtained was $87.8 \pm 2.3\%$. A similar implementation was presented in [6] to obtain biometric information of ECG signals. In this work, an accuracy of 99.48% was achieved, increasing the reliability of the BoF method in bio-medical signal classification.

There are works that implement modifications over the classic bag of features. In [15], different data representation for the bag of patterns were proposed. The signals were pre-processed by a technique named Symbolic Aggregate approXimation (SAX) presented on [16] which converted the signal in a text sequence. The first BoP proposal was called Multivariate Bag of Patterns and it was able to capture the relationship between time series over the time, this method created multi-variate words which represented the time series in time intervals. Another method was called Stacked Bag of Patterns where, unlike the latter method, each signal was treated as an individual BoP instance and then concatenated into a single one. The last method used Adapted Natural Language Processing Techniques on the two previous methods, document processing techniques were applied on the SAX signal, Term Frequency (TF), Inverse Document Frequency (IDF), Inverse Frequency (IF) and the combination IDF-IF.

In [17], an improvement on the clustering step was presented, named as *Bag of Super-Features*. It consists on generate clusters for each class on the dataset and later merge them. The underlying idea was that disregarding the labels in clustering step can lead to mitigation of significant differences [18].

In an attempt to classify long-term information, [19] proposed a Dual Layer Bag of Frames Model (DL-BoF), applied on music genre classification. This method applied a Bag of Frames with FFT characterization over a set of signals, later on, it created a second layer BoF, this was achieved by applying a Bag of Histograms Aggregation. Then a first layer dictionary was trained from FFT spectrograms and the second layer from histograms.

Temporal information is discarded when the BoW representation is applied since the extracted features are transformed into a histogram representation. Therefore, [20] proposed a temporal BoW which consists on divide a signal and then to apply the BoW to each segment, obtaining a set of histograms from each signal segment. Then the histograms must be combined to create a new instance.

An approach of EEG classification by means of BoW in complex environments is aimed in [7]. The proposed BoW was able to discriminate among two different tasks loads. Essentially, the proposed method was composed of three steps; identification of discriminative features, dictionary construction, and BoW modeling. The feature extraction was based on Wavelet decompositions and Independent Component Analysis. Further, a dictionary was generated by applying a k-means algorithm. The results showed that the BoW was able to discriminate two different tasks. Moreover, by analyzing the results, it was concluded that obtained patterns may vary among subjects.

2.3. Transfer learning in BCIs

Formally, transfer learning is defined in [21, 22] as follows. Given a source domain D_S and learning task T_S a target domain D_T and a learning task T_T, transfer learning aims to improve the learning of the target predictive function $f_T(\cdot)$ in D_T using the knowledge in D_S and T_S, where $D_S \neq D_T$, or $T_S \neq T_T$.

Moreover, following the descriptions in [21], the proposed method is categorized as Inductive Transfer Learning. In this case, the target task is different from the source task, no matter when their domains are the same or not. Also, there are available labeled data in the source domain, and transferring aims to achieve a high performance in the target task by using the source data. It is also mentioned that this approach allows the feature representation transfer, this is to find a feature representation that reduces the difference between the source and the target domains/tasks as well as the classification and regression error.

Due to the variability in brain signals across subjects, or even for the same subject [23], transfer learning has an area of opportunity in BCIs. Most of the transferring approaches are based on Common Spatial Patterns (CSP), assuming that there exists a set of invariant filters across sessions or subjects [24].

A different approach on CSP [25], proposed a combination of the CSP covariance matrices of different subjects for binary motor imagery classification. Moreover, a search of similarity among subjects was proposed by measuring the divergence between data distributions.

On the other hand, [26] has merged transfer learning with active learning for classification of visually evoked potentials in EEG. Active learning allows the selection of the most relevant data. However, the estimation of the most informative data is still an issue.

A deeper analysis of the signal was performed in [27] for subjects transfer learning, where T1-weighted Magnetic Resonance Imaging (MRI) is considered in addition to EEG recordings. The underlying idea is that training a model using the anatomical data may compensate the variability in electrode positioning and head morphology and, in consequence, improve the transferring. For this purpose, both EEG and MRI data were processed with Boundary Element Modelling and Minimum-Norm Estimate respectively to define the source activity.

In [28], in addition to transfer learning, the use of a low cost and user-friendly EEG device is analyzed. Moreover, the feature extraction methods are based in frequency bands extraction, and also the raw signal was analyzed. Following this, the analysis of the EEG raw signal, i.e. without frequency data extraction, is also proposed in this work.

There is an area of interest in the search of common brain activity patterns across subjects. Nevertheless, some approaches do not assume that a similar pattern between subjects exists. It is the case of [29], where a classifier is built for each subject and the transferring is performed by adjusting these classifiers for a new user.

In [30], a spatial filter bank generation per subject is proposed, such filters are assembled following selection criteria for motor imagery classification. The subject transferring is proposed by selecting the filter training sets of subjects whose characteristics are similar to the target user. A drawback of this method is the definition of the relevant features for different subjects.

2.4. Summary

Previous works in imagined speech have proposed different approaches as vowels and syllables analysis. Despite signal analysis of this approaches have a high discrimination performance, its main drawback is the limitation in the available vocabulary. The intuitive extension of vowels and syllables is the use of words, this allows to take advantage of the vocabulary context. For this reason, this approach is addressed in this work.

Although several works have explored the BoF representation some of the limitations are explored in this work. It is important to highlight that imagined speech, is a mental task which requires a higher level of abstraction in comparison to other tasks as motor imagery [31]. Due to that most of the previous works are focused in binary classification, a multi-class classification is considered in this work. Another consideration is the use of a simpler feature extraction based on frequency filtering and signals micro-volt values to generate the BoF.

Finally, most of the transferring approaches in BCIs are oriented to calibration time suppression, that is the transferring of a model without requiring data from the new subject or task. Two categories are proposed by [32]. First, a pooled design which optimizes a single model on the combined data of multiple users. Besides, an ensemble design, in which a model is optimized for each user and combined afterward. Following these concepts, this work fits in the definition of a pooled design. By means of transfer learning, the possibility to re-utilize a generated imagined speech codebook to model new imagined words for each subject is proposed, this can be seen as an intra-subject transfer learning task. BCI transferring is commonly focused on inter-subject variability, i.e. to extend a generated model to a new subject. The use of previously learned features to model instances of a new class is a specific case of transfer learning named as feature representation transfer.

3. Method

The Fig. 1 represents the general method flowchart, which is based on the work presented in [5, 6]. This method was applied individually to each subject.

3.1. Dataset

An imagined speech data set was recorded in [8], which is composed of the EEG signals of 27 native Spanish speaking subjects, registered through the Emotiv EPOC headset, which has 14 channels and a sampling frequency of 128 Hz. The data consist of 5 Spanish words (i.e. ”arriba”, ”abajo”, ”izquierda”,

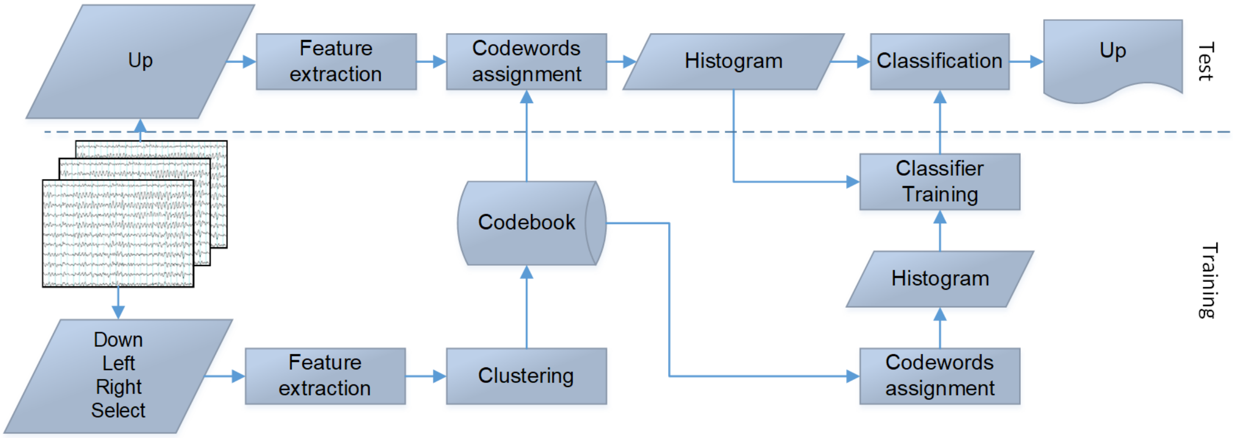

Figure 1: General method flowchart

"derecha", "seleccionar"; translated to English as "up", "down", "left", "right", "select" respectively) with 33 epochs each one and rest periods between them.

Data were processed with Common Average Reference (CAR). Also, a low-pass filter to reduce the noise was applied, such filter is an infinite impulse response Butterworth filter with a stop-band frequency of 50 Hz and a pass-band frequency of 40 Hz. No additional signal processing was applied, and the micro-volt signal values were used as features.

3.2. Feature extraction

Each imagined word EEG epoch, i.e. a word repetition, w_i for the s_j subject can be seen as a n by 14 matrix X_{w_i,s_j}, where 14 is the number of channels and n is the recording time for each epoch.

$$X_{w_i,s_j} = \begin{bmatrix} x_{1,1} & x_{1,2} & \dots & x_{1,n} \\ x_{2,1} & x_{2,2} & \dots & x_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{14,1} & x_{14,2} & \dots & x_{14,n} \end{bmatrix}$$

From the $X_{wi,sj}$ matrix, a transformation was applied to keep the spatial information of the signal as a new representation, this allowed a pattern detection from different areas of the brain and their activation during the imagined speech. In matrix y representation, the micro-volt values of the signal were taken as features.

Then the matrix y instances were generated by taking samples from all channels at the same time instant and concatenating them into a single vector. Resulting in n instances per epoch.

$$y = \begin{bmatrix} [x_{1,1} \; x_{2,1} \; \dots \; x_{14,1}] \\ [x_{1,2} \; x_{2,2} \; \dots \; x_{14,2}] \\ \vdots \\ [x_{1,n} \; x_{2,n} \; \dots \; x_{14,n}] \end{bmatrix}$$

3.3. Codebook generation

The previously extracted features were used to generate a codebook through a clustering method, i.e. k-means, the objective was to obtain the most representative features from the data. As in [17], the clustering method is applied per each class (i.e. each imagined word). Then, for each class, k clusters are obtained and receive the name of codewords. Later the resulting k prototypes were concatenated into a single codebook [18].

A drawback by using k-means is that the number of clusters must be defined *a priori*. To define the number of clusters, knowledge of the problem and the data is essential. Due to the few reported works in imagined speech analysis, this issue was overcome by applying a genetic algorithm in order to find an appropriate number of clusters based on classification accuracy [2]. The clusters were obtained considering the classification accuracy as the objective function. Thus, the mean accuracy of every subject considering all the classes was used in order to obtain a unique cluster number. Such experiments obtained a cluster number k equal to 250.

These numbers will be fixed for the following experiments. Once the codebook was generated the next step was to replace every instance in matrix y, which generated the codebook, with one of the k codewords. The result was a sequence of codewords over the original data. To choose the codeword which replaced each instance, a similarity measure was applied, i.e., the Euclidean distance.

3.4. Histogram generation

At this step, the original signals became sequences of codewords. Then, the occurrence of the codewords was counted in each word epoch, for this task a convenient representation was a histogram. The result was a set of histograms, each of them representing an epoch of the different imagined words from each subject.

3.5. Classification

Once the signals were transformed into a set of histograms, the classes, i.e. imagined words, were associated with the corresponding set of histograms which represented them. Thus, each histogram was considered a classification instance. In order to analyze these histograms, a Naive Bayes classifier was applied. For this purpose, 75% of the word epochs were used to generate the codebook and train the classifier, and the remaining 25% were reserved for test purposes. Moreover, due to the random properties of the k-means, the method was applied in a 10-cross-fold validation of the partitions.

3.6. Transfer learning

For testing the transfer learning, the codebook generation was calculated with only four words from the dataset. Later on, the instances of a new imagined word were replaced using a codebook generated with these previous words. Thus, a set of histograms were also generated from the new word, these histograms were associated with this new class and finally, were merged with the previous classes histograms into a classifier training.

This set of histograms was used to find patterns able to discriminate the new class using codewords generated by different imagined words. Thus, the aim was to verify whether the new imagined word could be represented and discriminated with a previously generated codebook.

3.7. Calibration

The objective of the calibration analysis was to observe the classification performance by adding different amounts of instances of the new class in the codebook generation. A model able to discriminate without data of the new class is preferable. Nevertheless, in some cases, small amounts of data is required at the model generation to improve the discrimination of a new class.

4. Results

4.1. Transfer learning

First, the baseline result of the method was obtained using the entire vocabulary, i.e. five words are used in the codebook generation, achieving an average accuracy of $65.65\% \pm 13.39$. In Fig. 2 a comparison of the proposed representation and [8] is presented.

The average accuracy when transfer learning is applied to the “up” word was $58.74\% \pm 13.39$. Moreover, when transfer learning is applied to the “down” word, the average accuracy was $61.38\% \pm 12.47$. It must be highlighted that transfer learning results were obtained without a calibration process.

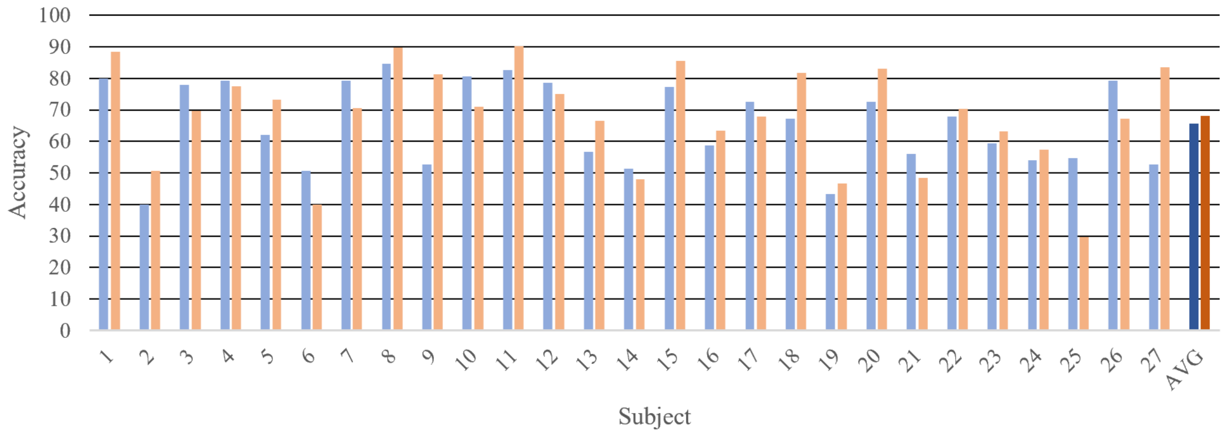

Figure 2: Accuracy results comparison without transfer learning (proposed method in blue, [8] in orange)

To contrast the transfer learning behavior, a baseline confusion matrix is showed in Table 1. The following confusion matrices were generated by averaging the confusion matrices from all subjects, and they are presented as global percentages for an easy interpretation. Thus, the confusion matrix in Table 1 corresponds to the results obtained in Fig. 1.

Table 1: Baseline confusion matrix.

	Up	Down	Left	Right	Select
Up	75.93	9.26	3.21	6.42	5.19
Down	8.89	61.48	7.78	14.57	7.28
Left	2.84	6.91	60.00	15.31	14.94
Right	3.83	12.96	13.33	61.36	8.52
Select	3.83	7.04	13.95	5.68	69.51

The confusion matrix of ”up” word transferring is presented in Table 2.

Table 2: “Up” class transferring confusion matrix.

	Up	Down	Left	Right	Select
Up	67.41	12.10	6.79	9.01	4.69
Down	15.19	56.17	8.52	13.21	6.91
Left	10.00	14.20	50.12	12.59	13.09
Right	10.49	20.37	10.62	51.23	7.28
Select	8.40	10.00	8.52	4.32	68.77

Also, in Fig. 3 the results per subject for the class ”Up” transferring are presented.

In addition, Table 3 corresponds to the “down” word transferring classification matrix, this table shows a different behavior from the transferred class. The accuracy obtained from “up” word decreases 8.52 in comparison to the baseline. Otherwise, the “down” word accuracy increased in 2.35.

Also, the average histogram from all subjects was obtained to calculate which codewords are used by the new class. This analysis should complement the confusion matrix analysis, comparing the confusion among classes and the percentage of codewords used. Table 4 summarizes the codewords percentage used in the transferred classes by averaging the results from all subjects.

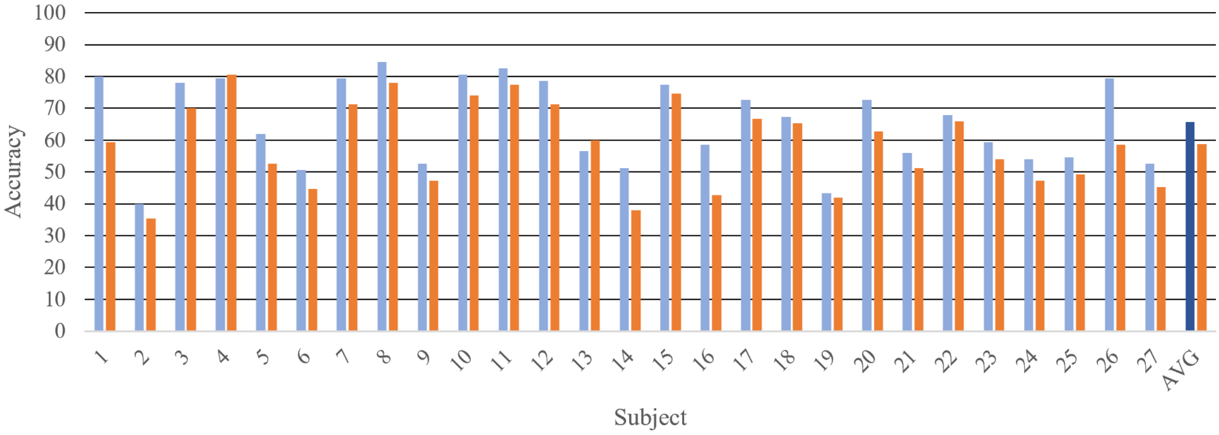

Figure 3: Transfer learning results for the word “up” (baseline in blue, transfer results in orange)

Table 3: “Down” class transferring confusion matrix.

	Up	Down	Left	Right	Select
Up	70.99	14.69	6.17	4.94	3.21
Down	9.88	58.52	13.09	13.58	4.94
Left	3.09	11.36	58.40	13.21	13.95
Right	3.70	18.27	20.00	52.59	5.43
Select	2.72	8.27	19.14	3.46	66.42

4.2. Calibration

To improve the classification results, a small epochs amount of the new word were used for the codebook generation. Then, the same number of instances used in the codebook generation were used for the classifier training. The class unbalancing presented by this calibration was not considered.

Table 5 shows the accuracies for the transfer learning approach, by using different amounts of epochs for the “up” word transferring.

Table 6 shows the accuracies from the “down” word in a transfer learning approach using different amounts of its epochs in the codebook generation step.

Both last two tables show the accuracy of the transferred classes and the total accuracy of the five classes, averaging the result of all the subjects.

5. Discussion

The baseline accuracy, when no transfer learning is applied is $65.65\% \pm 13.39$, this result has no significant effect with the baseline work in [8] with $[F(1, 52) = 0.4, p = 0.5323]$, according to a one-way analysis of variance of the subjects results. When applying transfer learning, a mean accuracy of 58.74 ± 13.39 was obtained for the “up” word, this is an accuracy decreasing of 6.91. When “down” word is transferred a mean accuracy of 61.38 ± 12.47 was achieved, this is an accuracy decreasing of 4.27. The transferring shows a slight decreasing compared to the baseline accuracy. Moreover, a one-way analysis of variance of the obtained results was conducted to compare the effect between the transfer learning approach of two words and the baseline method. There was no significant effect $[F(1, 52) = 3.6, p = 0.0634]$ for the ”up” word and $[F(1, 52) = 1.47, p = 0.2305]$ for the ”down” word.

In Table 2, the “up” word transferring confusion matrix shows a confusion between this word and “down” word. Moreover, as expected, most of the words increased its confusion with the word ”up” due to the results in Table 4 which showed that this new word is represented by instances of every word.

Table 4: Codewords distribution percentages to represent the transferred classes.

	Up	Down	Left	Right	Select
Up	-	22.45	23.26	26.02	27.56
Down	18.56	-	24.18	27.23	29.31

Table 5: "Up" class accuracies using epochs of this class in the codebook generation.

Epochs number	"Up" accuracy	Total accuracy
0	67.41 ± 19.22	58.74 ± 13.39
1	19.75 ± 18.00	56.77 ± 11.70
3	30.99 ± 19.59	61.83 ± 10.96
5	39.14 ± 21.83	61.60 ± 12.78
8	47.04 ± 24.01	62.79 ± 13.03
10	51.36 ± 23.47	62.79 ± 13.03

The "down" word shows a different behavior, results in Table 3, showed an accuracy of $61.38\% \pm 12.47$. It is interesting to highlight that the results do not correspond to the codewords distribution showed in Table 4. The "down" word codification included more codewords of the "select" word. Thus, a higher confusion between the words "down" and "select" was expected. Nevertheless, the "down" transferring confusion matrix shows a higher confusion with the "right" word.

In Table 5, the accuracy when calibration data is added to the "up" word transferring is presented. As it was expected, the accuracy increased as more calibration instances were included. Nevertheless, a higher performance was achieved when there was no information about the class. Otherwise, Table 6 shows the accuracy when the "down" word was transferred. It can be noticed that the transferring of this word shows a similar behavior as the "up" word.

The performance decrement when calibration is applied could be attributed to three causes: (1) the use of few data to represent the transferred class, (2) the use of non-representative data of the new class, and (3) the class unbalancing at the classifier training. In a practical application, the calibration instances can be seen as instances of a new word that a user wants to learn. By a deeper analysis, these instances could be exploited to improve the knowledge of the new imagined word and increase the recognition using a few calibration instances.

6. Conclusions

The proposed method showed to be capable to extend an imagined speech vocabulary with a slight accuracy decrement for a set of subjects. By excluding the data of the transferred word from the generation step, it obtained similar accuracy results as if the data was included. Nevertheless, the use of few calibration data did not help to increase the classification accuracies. For practical applications, this calibration step must use the less information possible from the transferred word. It must be considered that, if the calibration data are not representative of the imagined word, the codebook will not represent correctly such word and the classification performance will decay.

It is possible to add more than one word to the model. Nevertheless, it is expected that the classification performance decrement. Further experiments may analyze the method behavior when more words are added, considering different combinations of the added words and calibration steps.

The analysis of the codewords distribution showed that the number of codewords used from other imagined words is not correlated to the confusion among classes. Thus, a deeper analysis must be performed to explain these results properly.

It is also interesting to highlight that the feature extraction step takes only into account the signal microvolt values. Hence, the impact of noise in the filtered signals must be explored. Additionally, in future experiments, the frequency information of the signal could be taken into account to search for patterns

Table 6: "Down" class accuracies using epochs of this class in the codebook generation.

Epochs number	"Down" accuracy	Total accuracy
0	58.52 ± 24.62	61.38 ± 12.47
1	15.19 ± 13.34	58.07 ± 11.08
3	22.96 ± 16.44	60.59 ± 10.95
5	27.90 ± 16.70	61.83 ± 10.96
8	33.70 ± 22.91	62.59 ± 12.92
10	38.52 ± 22.73	62.74 ± 13.06

related to the brain activity frequency bands. Nevertheless, the omission of frequency features extraction step of the signal makes the method more suitable for a real-time BCI.

7. Acknowledgments

The present work was partially supported by CONACyT (scholarship 401887). Also, the authors thank the support of the Italian Foreign Affairs and Cooperation Ministry, and the International Cooperation for Development Mexican Agency for the project MX14MO06.

References

[1] G. A. Pressel Coretto, I. E. Gareis, H. L. Rufiner, Open access database of eeg signals recorded during imagined speech, Proc. SPIE 10160 (2017) 1016002–1016002–11. doi:10.1117/12.2255697.
URL http://dx.doi.org/10.1117/12.2255697

[2] J. S. García-Salinas, L. Villaseñor-Pineda, C. Reyes-García, A. A. Torres-García, Selección de parámetros en el enfoque de bolsa de características para clasificación de habla imaginada en electroencefalogramas, Research in Computing Science 1 (140) (2017) 123–133.

[3] P. Suppes, Z.-L. Lu, B. Han, Brain wave r ecognition of words, Psychology 94 (1997) 14965–14969.

[4] M. Salama, H. Lashin, T. Gamal, Recognition of unspoken words using electrode electroencephalograhic signals, COGNITIVE 2014 : The Sixth International Conference on Advanced Cognitive Technologies and Applications (2014) 51–55.

[5] J. Wang, P. Liu, M. F. She, S. Nahavandi, A. Kouzani, Bag-of-words representation for biomedical time series classification, Biomedical Signal Processing and Control 8 (6) (2013) 634–644. doi:https://doi.org/10.1016/j.bspc.2013.06.004.

[6] J. Wang, M. She, S. Nahavandi, A. Kouzani, Human identification from ecg signals via sparse representation of local segments, IEEE Signal Processing Letters 20 (10) (2013) 937–940.

[7] L. M. Meriño, J. Meng, S. Gordon, B. J. Lance, T. Johnson, V. Paul, K. Robbins, J. M. Vettel, Y. Huang, A bag-of-words model for task-load prediction from eeg in complex environments, Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on (2013) 1227–1231.

[8] A. A. Torres-García, C. A. Reyes-García, L. Villaseñor-Pineda, G. García-Aguilar, Implementing a fuzzy inference system in a multi-objective eeg channel selection model for imagined speech classification, Expert Systems with Applications 59 (2016) 1–12.

[9] C. H. Nguyen, G. Karavas, P. Artemiadis, Inferring imagined speech using EEG signals: a new approach using Riemannian Manifold features, Journal of Neural Engineeringdoi:10.1088/1741-2552/aa8235.
URL http://iopscience.iop.org/10.1088/1741-2552/aa8235

[10] S. Zhao, F. Rudzicz, Classifying phonological categories in imagined and articulated speech, 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (2015) 992–996doi:10.1109/ICASSP.2015.7178118.
URL http://ieeexplore.ieee.org/document/7178118/

[11] A. A. Torres-García, C. A. Reyes-García, L. V.-P. L., J. Ramirez, Analisis de señales electroencefalograficas para la clasificacion de habla imaginada, Revista mexicana de ingeniería biomedica 34 (2013) 23 – 39.

[12] M. Baydogan, G. Runger, E. Tuv, A bag-of-features framework to classify time series, IEEE Transactions on Pattern Analysis and Machine Intelligence 35 (11) (2013) 2796–2802. doi:10.1109/TPAMI.2013.72.

[13] J. Lin, Y. Li, Finding structural similarity in time series data using bag-of-patterns representation, Scientific and Statistical Database Management: 21st International Conference, SSDBM 2009 New Orleans, LA, USA, June 2-4, 2009 Proceedings (2009) 461 – 477.

[14] J. Wang, P. Liu, M. F. She, S. Nahavandi, A. Kouzani, Bag-of-words representation for biomedical time series classification, Biomedical Signal Processing and Control 8 (6) (2013) 634 – 644. doi:http://dx.doi.org/10.1016/j.bspc.2013.06.004.
URL http://www.sciencedirect.com/science/article/pii/S174680941300089X

[15] P. Ordonez, T. Armstrong, T. Oates, J. Fackler, Using modified multivariate bag-of-words models to classify physiological data, in: 2011 IEEE 11th International Conference on Data Mining Workshops, 2011, pp. 534–539. doi:10.1109/ICDMW.2011.174.

[16] J. Lin, E. Keogh, L. Wei, S. Lonardi, Experiencing sax: a novel symbolic representation of time series, Data Mining and Knowledge Discovery 15 (2) (2007) 107–144. doi:10.1007/s10618-007-0064-z.
URL http://dx.doi.org/10.1007/s10618-007-0064-z
[17] A. Plinge, R. Grzeszick, G. A. Fink, A bag-of-features approach to acoustic event detection, in: 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2014, pp. 3704–3708. doi:10.1109/ICASSP.2014.6854293.
[18] S. Lazebnik, M. Raginsky, Supervised learning of quantizer codebooks by information loss minimization, IEEE Transactions on Pattern Analysis and Machine Intelligence 31 (7) (2009) 1294–1309. doi:10.1109/TPAMI.2008.138.
[19] C. C. M. Yeh, L. Su, Y. H. Yang, Dual-layer bag-of-frames model for music genre classification, in: 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, 2013, pp. 246–250. doi:10.1109/ICASSP.2013.6637646.
[20] Z.-W. Gui, Y.-R. Yeh, Time series classification with temporal bag-of-words model, in: Technologies and Applications of Artificial Intelligence, Springer, 2014, pp. 145–153.
[21] S. J. Pan, Q. Yang, A survey on transfer learning, IEEE Transactions on Knowledge and Data Engineering 22 (10) (2010) 1345–1359. doi:10.1109/TKDE.2009.191.
[22] H. He, D. Wu, Transfer learning enhanced common spatial pattern filtering for brain computer interfaces (bcis): Overview and a new approach, in: International Conference on Neural Information Processing, Springer, 2017, pp. 811–821.
[23] H. Morioka, A. Kanemura, J. ichiro Hirayama, M. Shikauchi, T. Ogawa, S. Ikeda, M. Kawanabe, S. Ishii, Learning a common dictionary for subject-transfer decoding with resting calibration, NeuroImage 111 (2015) 167–178. doi:10.1016/j.neuroimage.2015.02.015.
[24] V. Jayaram, M. Alamgir, Y. Altun, B. Scholkopf, M. Grosse-Wentrup, Transfer Learning in Brain-Computer Interfaces, IEEE Computational Intelligence Magazine 11 (1) (2016) 20–31. arXiv:1512.00296, doi:10.1109/MCI.2015.2501545.
[25] H. Kang, Y. Nam, S. Choi, Composite common spatial pattern for subject-to-subject transfer, IEEE Signal Processing Letters 16 (8) (2009) 683–686. doi:10.1109/LSP.2009.2022557.
[26] D. Wu, B. Lance, V. Lawhern, Transfer learning and active transfer learning for reducing calibration data in single-trial classification of visually-evoked potentials, in: Systems, Man and Cybernetics (SMC), 2014 IEEE International Conference on, IEEE, 2014, pp. 2801–2807.
[27] M. Wronkiewicz, E. Larson, A. K. C. Lee, Leveraging anatomical information to improve transfer learning in brain computer interfaces, Journal of Neural Engineering 12 (4) (2015) 046027.
[28] G. Panagopoulos, Multi-Task Learning for Commercial Brain Computer Interfaces, 17th International Conference on Bioinformatics and Bioengineering (2017) 86–93doi:10.1109/BIBE.2017.00022.
[29] S. Dalhoumi, G. Dray, J. Montmain, Knowledge transfer for reducing calibration time in brain-computer interfacing, in: Tools with Artificial Intelligence (ICTAI), 2014 IEEE 26th International Conference on, IEEE, 2014, pp. 634–639.
[30] W. Tu, S. Sun, A subject transfer framework for EEG classification, Neurocomputing 82 (2012) 109–116. doi:10.1016/j.neucom.2011.10.024.
[31] K. Brigham, B. V. K. V. Kumar, Imagined Speech Classification with EEG Signals for Silent Communication: A Preliminary Investigation into Synthetic Telepathy, 2010 4th International Conference on Bioinformatics and Biomedical Engineering (2010) 1–4doi:10.1109/ICBBE.2010.5515807.
URL http://ieeexplore.ieee.org/document/5515807/
[32] F. Lotte, Signal processing approaches to minimize or suppress calibration time in oscillatory activity-based brain-computer interfaces, Proceedings of the IEEEdoi:10.1109/JPROC.2015.2404941.