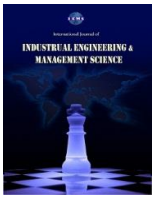
University of Hormozgan

International Journal of Industrial Engineering & Management Science

journal homepage: www.ijiems.com




# Imagined Speech Classification Accuracy and the Signal Acquisition Procedure

Gholam Reza Mohammad Khani [a,*], Mohammad Reza Asghari Bejestani [a]

[a] *Electrical & IT department, Iranian Research Organization for Science and Technology (IROST), Tehran, Iran*

ARTICLE INFO

*Article history:*
Received: 2023-05-29
Received in revised form: 2023-06-24
Accepted: 2023-07-05

*Keywords:*
Imagined speech
EEG Signal Acquisition
Brain Computer Interface (BCI)

ABSTRACT

Imagined speech recognition is one of the most interesting approaches to BCI development. A lot of works have been done in this area. Many different experiments have been designed and hundreds of combinations of feature extraction methods and classifiers have been examined. Reported classification accuracies range from the chance level to more than 90%. Based on non-stationary nature of brain signals, we have introduced 3 classification modes according to time difference in inter and intra-class samples. The modes can explain the diversity of reported results and predict the range of expected classification accuracies from the brain signal accusation procedure. In this paper, a few samples are illustrated by inspecting results of some previous works. It has been shown that, in current state of art researches on imagined word classification, if signal accusation schema falls in the mixed time mode, the accuracy can reach to more than 90 percent's, but for more realistic short and long time modes, it's hard to attain good results.

## 1. Introduction

A Brain-Computer interface (BCI) is a system that translates brain signals into some other kinds of outputs [1]. BCI systems generally consist of four steps: Signal acquisition, Preprocessing, Feature extraction and Classification.

In Signal acquisition step, some signals related to the brain activities are captured and bring into the system. Preprocessing involves noise reduction, signal segmentation and similar tasks to highlight and bring into hand the most useful signal parts. Feature extraction is used to extract those signal properties to be used in classification step. Feature extraction usually reduces processed data size by removing none useful information in the signal. Finally, classification is where we decide that the input signal corresponds to which of output classes. Artificial Neural Networks (ANN), Support Vector Machines (SVM), K-Nearest Neighbors (KNN), Deep Learned machines, etc. are examples of many available methods and algorithms (classifiers) to map the extracted signal features to a purposed task.

The focus of this paper is on signal acquisition phase. Different signal types and capturing devices may be used. In magneto¬encephalo-graphy (MEG) magnetic fields produced by neuron activities are used. Electrical fields are measured by electro¬encephalo¬graphy (EEG) and Electro¬cortico¬graphy (ECoG). In EEG, electrodes are placed on outer side of the skull, while ECoG uses electrodes placed directly on the exposed surface of the brain. Functional near-infrared spectroscopy (fNIRS), measures brain activity by using near-infrared light to estimate cortical hemodynamic activity which occur in response to neural activity [2]. Most of recent works on BCI developments use EEG signals as the input of their systems because of its ease of use, low cost, high temporal resolution and safety [3].

The type of brain activity is also different among BCIs. In motor imagery experiments, the participants are asked to perform a motor imagination task such as moving a finger, an arm or legs [4,5]. For image perception tasks, the subjects

* Corresponding author.
*E-mail address:* mhmdreza46@yahoo.com

concentrate on watching some displayed pictures, for example, simple geometric shapes (rectangle, circle, triangle, etc.), real pictures (persons, animals, planets, objects, etc.), or even written words and letters.

Imagined speech recognition (aka Silent Speech) is one of the commonly used approaches for implementing BCI systems. In speech imagination, the participants imagine the pronunciation of a particular vowel [6-8], syllable [9-11] or word [12,13] in some defined time intervals. EEG signal during these intervals are processed to determine the imagined word [11,12,14].

On the other hand, BCI systems may be on-line or off-line. An on-line system, performs all above 4 steps at the same time that the subject performs the required task. For example, in a robotic arm control BCI, the subjects brain signal needs to be processed on-line (perhaps in real time) in order to generat appropriate commands for movement of the arm. This type of BCIs are preferred for real life applications. In off-line systems, the required signals are recorded and then processed and classified in off-line. This type of systems is used for development of processing algorithms.

In [15] we presented a work on off-line imagined Farsi words recognition using EEG signals. There, we classified the imagination signals of 6 Persian words and the scilence. The word were Farsi equvalents of “LEFT”, “RIGHT”, “UP”, “DOWN”, “YES” and “NO”, which can be used for moving an object (mouse cursor, a robotic arm, wheelchair etc.) or to answer Yes/No questions. A binary SVM machine was trained to classifiy between two specific words or a word and the scilence. The feature sets were normalized spectrum of 19 EEG channels placed according to 10-20 standard.

The bank of SVM machines were used for different 2, 3 or 7 class classification problems. Besides the classification accuracies, one of the main findings of the research was the effect of signal acquisition procedure on the classification accuracy. In this paper, we explain this effect and illustrate its trace in results of some previous works.

## 2. EEG Signal Acquisition Procedures

Working on brain signals usually needs an experiment which guides the subject to perform and concentrate on a specific task. In case of off-line imagined word classification, the task is imagining one of predefined words (item). In order to record EEG signal corresponding to an item (a trial), the simple procedure of Figure 1 may be used. Stimulation is a signal (visual, auditory, etc.) presented to the subject to inform him/her to perform the required task. After a small settling time, subject starts to imagine the requested word and finally a small rest time is given to the subject to relax for starting another trial. The same procedure repeats several times for each item. Usually only a portion of the brain signal during imagination of an item is considered as a sample for that item.

For the purpose of this research, types of stimulation, number and types of the items and even the timing of trials are not important. We will analyze the accuracy of classification processes based on the time difference between their trial samples.

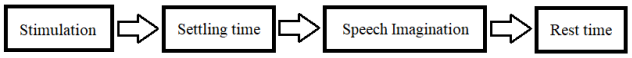



**Figure 1**. Time sequence of a single trial

## 3. Classification Modes

Statistics of EEG signals is known to change during the time. Therefore, the differences between the times of recording EEG signals (TDs), can directly affect the performance of classification. Based on TDs of inter and intra-class samples, three modes of classification were distinguished in [15]:

a) Long-time classification where the TDs are typically long for both the samples of instances of a single class and the all samples in all classes. This mode happens when signal samples from a relatively long time span are used in classification, therefore the deference in recording time of samples of the same class, may be from a few seconds to several months, i.e. the time during which the EEG signals were recorded. The same is true for intra-class TDs. If the classifier has a training phase, this may apply to train and test samples as well. A common case of this mode are most on-line exprements where a pre-trined classifier, which is trained with previous old samples, is used to classify newer samples.

b) Short-time classification when the TDs are typically short for both inter and intra class samples. This mode happens when signal samples of all classes are recorded in a relatively short time so that recording time of all samples are near. For example, in some online processing seassions with simultaneous train and test procedures, the classification is more likely to be in this mode, because all train and test signals have been randomly distributed in a relatively small time periode.

c) Mixed-time classification which samples of every class are taken in a different time segment. Now, the inter-class DTs is short (as small as a few seconds) while the intra class time- the time between recordings of different classes- is typically large. For example, in expriments that all samples of every class are recorded in a different batch (a single periode of time), separated from the other classes. On-line classification may rarely fall in this mode.

## 4. Results and Discussion

The classification accuracies we found in our research [15, 16] are summarized in Table 1. The accuracies are nearly perfect for mixed-time mode, but decrease significantly for short-time and fall very close to chance in long-time mode. Various papers published for similar researches in the field of imagined speech recognition, have also reported different accuracies varying from near chance level up to more than 90%. We believe that the remarkable difference in classification accuracies in these

researches, is mostly due to their signal acquisition procedure which yields to different classification modes. There are so many papers which their result could be analyzed in this way. Here we present a few examples.

Table 1. Summary of average accuracies in all classification types and modes in [15]

| Classification Mode | Classification Accuracy ( mean ± standard deviation) | | | |
|---|---|---|---|---|
| | Word-Silence | Word-Word | Word-Word-Silence | 6 Words + Silence |
| Short-time | 95.7± 8.3 | 74.5± 20.4 | 78.9± 13.9 | 55.2± 8.8 |
| Long-time | 57.5± 7.2 | 59.5± 7.5 | 39.5± 5.4 | 32.0± 2.1 |
| Mixed-time | 97.1± 5.1 | 96.4± 8.0 | 92.0± 11.1 | 88.2± 4.8 |

### 4.1. Chi et.al. 2011 [17]

One of the papers which has reported very similar results as our research is Chi et.al. 2011 [17]. It reports results of pairwise classification for imagination of five types of phonemes. Figure 2 shows a sample trial time course of their experiment.

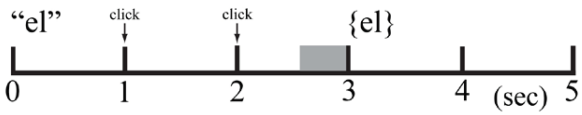



**Figure 2**. Trial time course illustrated for the -l phoneme [17]

The phoneme to be imagined was cued using a heard prompt presented at trial onset. This was followed by two audible clicks. The subject generated the cued phoneme in imagination during the 2 seconds long interval starting agter the last click. Data drawn from the 400 mSec period indicated by the gray rectangle and starting at time 2.6 sec were subject to classification analysis.

Twelve instances of each trial type were presented in block-randomized order during a single experimental session for a total of 144 trials per session. One, two, or three sessions were run on a single day, providing 24, 36 or 48 trials per phoneme, respectively, and 48, 72 or 96 trials per articulation class, respectively. The sessions were run in total for either four or three days.

Table 2 shows LDA classification performance in percent for 15 pairwise classifications of articulation class for 5 subjects A, B, C, D and E. The first number in each cell is the arithmetic average of the single-day classification performances. The second number is the performance found when classifying the entire dataset. So the first numbers are the results of a two class short-time mode classification and the second numbers correspond to long-time mode.

Comparing to our results, we can see that despite the difference in class types and signal timings, results are very similar: about 70% (71.7 ± 2.2) for two class short-time mode and 60% (63.0 ± 3.5) for two class long-time mode.

### 4.2. Al Saleh 2019

In [18], Al Saleh et.al have used five words -"Left", "Right", "Up", "Down" and "Select"- and performs two different experiments:

1) Mouse clicks (MC): Sixty trials (divided into two block of 40 and 20) were collected for every word. The participant made one mouse click immediately before and after each trial (i.e. the word imagination period). During the recording, the time between the end of one trial and the start of the next was decided by the participant and could be used as the rest time for the participant.

2) Specified time frame (TF): Forty trials for every word were collected as a block. The participants were given four seconds to imagine the pronunciation for each word followed by two seconds as the rest time between trials.

Again, despite the different trial designs, the recording procedure clearly yields to mixed-time mode classification because the TD between samples of the same class was very short (a few seconds) while for two different classes were more longer (at least a few minutes).

The paper reports classification results for 9 subjects and 4 classifiers (Table 3). As expected the accuracies are high and the best average result (87.4) matches our seven class mixed mode accuracy (88.2 %).

Table 2. Multi-class Classification performance [17] (see text for details)

| | Subject | Tongue | | Nasal | | Lips | | Fricative | | Relax | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Jaw** | A | 68.0 | *61.5* | *73.2* | 59.4 | *72.4* | *62.0* | *73.7* | *63.0* | *71.9* | *68.8* |
| | B | 66.9 | 58.6 | *70.1* | *62.2* | *69.6* | 60.7 | *71.8* | 60.4 | *73.7* | *68.8* |
| | C | *73.2* | *63.7* | 71.3 | *61.6* | *74.5* | 59.7 | *75.2* | 60.5 | *73.9* | *62.4* |
| | D | *71.7* | *60.7* | *69.8* | *60.7* | *73.0* | *60.7* | *74.8* | *61.2* | *70.8* | *61.9* |
| | E | *70.9* | *62.9* | *71.6* | *65.4* | *72.4* | *62.4* | 69.8 | *66.4* | *72.1* | *70.3* |
| **Tongue** | A | | | 67.7 | *65.6* | *75.3* | *62.2* | *74.0* | *66.4* | *72.1* | *69.5* |
| | B | | | 71.2 | *62.8* | *70.9* | 56.5 | 67.1 | 61.3 | *70.5* | *63.7* |
| | C | | | *74.0* | *64.7* | 71.3 | 58.9 | *75.8* | *62.1* | *70.6* | *61.8* |
| | D | | | 70.2 | *62.6* | *72.2* | 57.7 | *74.6* | 57.0 | 69.1 | *60.5* |
| | E | | | *76.0* | *66.1* | 69.7 | *65.0* | 70.6 | *68.2* | *72.3* | *63.6* |
| **Nasal** | A | | | | | *72.1* | 58.3 | *72.1* | *68.0* | *72.1* | *65.1* |
| | B | | | | | 69.3 | 58.6 | 72.1 | *64.0* | 70.6 | *67.0* |
| | C | | | | | *72.6* | *61.1* | *73.6* | 58.4 | *72.1* | *63.2* |
| | D | | | | | 66.4 | 59.1 | *71.5* | 59.8 | *70.3* | 59.1 |
| | E | | | | | *73.7* | *66.1* | *71.0* | *67.8* | *72.9* | *68.7* |
| **Lips** | A | | | | | | | 68.8 | *62.8* | *75.5* | *69.0* |
| | B | | | | | | | *69.6* | *62.2* | 68.1 | *62.8* |
| | C | | | | | | | *74.8* | *64.5* | *69.0* | 60.0 |
| | D | | | | | | | *73.1* | *60.3* | 69.5 | 57.7 |
| | E | | | | | | | *72.5* | *68.0* | 70.5 | *68.9* |
| **Fricative** | A | | | | | | | | | *75.5* | *69.0* |
| | B | | | | | | | | | *74.3* | *65.8* |
| | C | | | | | | | | | *71.0* | *64.5* |
| | D | | | | | | | | | *72.8* | 59.8 |
| | E | | | | | | | | | *71.5* | *66.6* |

Table 3. 5-Class average accuracy in mixed time mode classification of [18]

| Subject | Mouse Click | | | | Fixed Time Frame | | | |
|---|---|---|---|---|---|---|---|---|
| | [SVM] | [NB] | [RF] | [LDA] | [SVM] | [NB] | **[RF]** | [LDA] |
| S1 | 68.8 | 73.1 | 87.2 | 58.7 | 61.3 | 74.3 | **86.4** | 49.7 |
| S2 | 41.8 | 52.9 | 57.1 | 45.5 | 68.8 | 82.9 | **84.4** | 67.3 |
| S3 | 50.3 | 64.1 | 69.8 | 55 | 60.8 | 72.4 | **88.9** | 58.3 |
| S4 | 61.3 | 78.9 | 79.3 | 53.9 | 68.3 | 91 | **98.5** | 74.3 |
| S5 | 37 | 44.4 | 54.6 | 33.8 | 55.4 | 76.9 | **80.4** | 51.8 |
| S6 | 67.3 | 53 | 70.4 | 51.3 | 87.4 | 82.9 | **93.9** | 76.5 |
| S7 | 48.6 | 54.5 | 60.9 | 46.1 | 68.4 | 66.9 | **83.5** | 54.9 |
| S8 | 50.2 | 67.2 | 72 | 46 | 83.9 | 95 | **97** | 83.9 |
| S9 | 49.8 | 67.2 | 73.1 | 56.6 | 40.2 | 59.8 | **73.8** | 40.2 |
| Average | 52.7 | 61.7 | 69.3 | 49.6 | 66 | 78 | **87.4** | 61.8 |

## 5. Conclusion

The first step in development of BCI systems is the signal acquisition. Based on non-stationary nature of brain signals, at least during imagined speech, we have shown that the designed procedure for this step has a major effect on experiment's final results. In the mixed-time mode, the expected classification accuracy would be very high, but for short and log-time modes, but for short and log-time modes it falls to lower values, even to the chance level. Unfortunately, for real life online applications only the long and short-time modes should be used which are more likely yield to moderate and poor results. Therefore, future experiments should avoid bulk sampling- where samples of each class are recorded in a few separated and continuous time periods- and focus on class samples which have distributed randomly on a short or long time interval.

## References

[1] Nijboer, F; Sellers, E; Mellinger, J; Jordan, M. A, Matuz,T; Furdea, A; Halder, S; Mochty, U; Krusienski, D.J; Vaughan, T. M, "A P300-based brain–computer interface for people with amyotrophic lateral sclerosis," Clinical neurophysiology, vol. 119, no. 8, pp. 1909-1916, 2008.

[2] M. Ferrar and V. Quaresima, "A brief review on the history of human functional near-infrared spectroscopy (fNIRS) development and fields of application," NeuroImage, vol. 63, no. 2, pp. 921-935, 2012.

[3] D. Lopez-Bernal, D. Balderas, P. Ponce and A. Molina, "A State-of-the-Art Review of EEG-Based Imagined Speech Decoding," Frontiers in Human Neuroscience, vol. 16:867281, 2022.

[4] L. Mondada, M. Karim and F. Mondada, "Electroencephalography as implicit communication channel

for proximal interaction between humans and robot swarms," Swarm Intelligence, vol. 10, no. 4, pp. 247-265, 2016.

[5] G. Pfurtscheller and C. Neuper, "Motor imagery and direct brain-computer communication," Proceedings of the IEEE, vol. 89, no. 7, pp. 1123-1134, 2001.

[6] DaSalla, C.S; Kambara, H; Koike, Y; Sato, M, "Spatial filtering and single-trial classification of EEG during vowel speech imagery," in Proceedings of the 3rd International Convention on Rehabilitation Engineering & Assistive Technology, 2009.

[7] C. 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.

[8] Idrees, B.M; Farooq, O, "Vowel classification using wavelet decomposition during speech imagery," in 2016 3rd International Conference on Signal Processing and Integrated Networks (SPIN), Noida, India, 2016.

[9] S. Deng, R. Srinivasan, T. Lappas and M. D'Zmura, "EEG classification of imagined syllable rhythm using Hilbert spectrum methods," Journal of neural engineering, vol. 7, no. 4, p. 046006, 2010.

[10] Kim, J; Lee, S.K; Lee, B, "EEG classification in a single-trial basis for vowel speech perception using multivariate empirical mode decomposition," Journal of neural engineering, vol. 11, no. 3, p. 036010, 2014.

[11] D'Zmura, M; Deng, S; Lappas, T; Thorpe, S; Srinivasan,R, "Toward EEG Sensing of Imagined Speech," in Human-Computer Interaction. New Trends, Berlin, Heidelberg, 2009.

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

[13] Saha, P; Abdul-Mageed, M; Fels, S, "SPEAK YOUR MIND! Towards Imagined Speech Recognition With Hierarchical Deep Learning," in Interspeech 2019, 20th Annual Conference of the International Speech Communication Association, Graz, Austria, 2019.

[14] D.-Y. Lee, M. Lee and S.-W. Lee, "Decoding Imagined Speech Based on Deep Metric Learning for Intuitive BCI Communication," IEEE transactions on neural systems and rehabilitation engineering: a publication of the IEEE Engineering in Medicine and Biology Society, vol. 29, pp. 1363-1374, 2021.

[15] M. R. Asghari Bejestani, G. R. Mohammad Khani, V. R. Nafisi and F. Darakeh, "EEG-Based Multiword Imagined Speech Classification for Persian Words," BioMed Research International, Vols. 2022, Article ID 8333084, 2022.

[16] M. R. Asghari Bejestani, G. R. Mohammadkhani, S. Gorgin, V. R. Nafisi and G. R. Farahani, "Classification of EEG Signals for Discrimination of Two Imagined Words," Journal of Signal and Data Processing(JSDP), vol. 17, no. 2, pp. 113-120, 2020.

[17] Chi, X; Hagedorna, J.B; Schoonovera, D; D'Zmuraa, M, "EEG-based discrimination of imagined speech phonemes," International Journal of Bioelectromagnetism, vol. 13, no. 4, pp. 201-206, 2011.

[18] AlSaleh, M; Moore, R; Christensen, H; Arvaneh, M, "Examining Temporal Variations in Recognizing Unspoken Words using EEG Signals," in 2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Miyazaki, Japan, 2019.