Recognition of Unspoken Words Using Electrode Electroencephalograhic Signals

May Salama, Loa'ay ElSherif, Haytham Lashin, Tarek Gamal
Computer Engineering Dept.
Faculty of Engineering at Shoubra, Benha University
108 Shoubra St. Cairo, Egypt
emails: {msalama@megacom-int.com, thefieryarroow@hotmail.com, haytham.lashin@yahoo.com, tarek_egypte@yahoo.com}

***Abstract*—In this paper, the use of electroencephalograhic signals in recognizing unspoken speech is investigated. The aim of the work is to recognize two words, namely, Yes and No by using a single electrode electroencephalograhic device. Yes and No were selected based on the fact that they are the basic answers of any disabled person. The single electrode device was chosen because of its ease of setup, adjustment, and light weight; hence, most suitable also for disabled. Several neural networks were trained with 7 subjects. Online and offline testing were carried out on male and female subjects in semi quiet environment. Average recognition results reached were 57% for online testing and 59% offline, whereas maximum value for offline was 68% and online 90%. The novelty of the work is using a single electrode device, as all previous work was done on multi-electrode devices.**

Keywords-electroencephalograhic signal; unspoken speech; recognition.

I. Introduction

Brain computer interface (BCI) is a communication between a person and a computer without physical movement. A signal emitted from the brain is measured, processed and interpreted into action on the computer. An electroencephalograhic (EEG) device records electrical signals from the brain using single or multiple electrodes. According to Graimann et al. [1], a BCI must have four components: 1- record activity directly from the brain, 2- must provide feedback to the user, 3- must do that in real time, 4- the user must choose to perform a mental task whenever a goal is needed to be achieved.

Recognizing unspoken speech is getting more attention due to its importance in many fields. One field is verbal disability of some people, while another is security field, where it is unsafe to speak in the presence of others. There are two approaches concerning recognition of unspoken speech; words and blocks (syllables). Wester and Shultz [2] implemented a system to recognize unspoken speech in five modalities using 20 electrode EEG device. The system used the blocks approach, as well as words approach. In continuation to Wester's work, Wester and Schultz [2] and Calliess [3] showed that the block wise presentation produced better results. He reached a 15.5% better than random guess, a best of 87.3%. Calliess [3] also raised the question whether the results of Wester and Schultz [2] were due to temporal brain artifacts. Two hypotheses were investigated by Porbadnigk [4]. First, Silent speech can be recognized based on EEG signals. Second, unspoken speech cannot be recognized based on EEG signals and that the good results reached by Wester and Schultz [2] were due to the temporal patterns that were recognized. The conclusion was that "It could be shown that except for the block mode which yielded an average recognition rate of 45.50%, all other modes had recognition rates at chance level". This was justified by the fact that block data would contain less noise and that the block mode made it easier to think about words in a consistent way. It was also concluded from work of Porbadnigk [4], that temporal artifacts are superimposed over the signal under investigation in block mode.

Some studies conducted by Wester and Schultz [2], Calliess [3], and Porbadnigk [4] were done in 2009, where a 16 channel EEG channels using 128 cap montage was used to recognize 5 words. Porbadnigk et al. [5] showed that the block mode yielded an average recognition rate of 45.5% and it dropped to chance level for other modes.

A research for recognizing unspoken 5 words using 21 subjects was carried out by Torres-García et al. [6]. These subsets were used to train four classifiers: Naïve Bayes (NB), Random Forests (RF), support vector machine (SVM), and Bagging-RF. The accuracy rates were above 20%.

The paper is organized in seven sections. After the introduction, overview of the system is presented in Section II. Section III shows feature extraction methods. Section IV discusses classification methods applied, while Sections V and VI investigate the testing results and their validation. Section VII concludes the paper and the future work.

II. System overview

A. Scope and limitations

The scope of words to be recognized is 2 arabic words "Yes" and "No" using the single electrode Neurosky Mindwave Mobile headset [7] to acquire brain signals. The device consists of a headset, an ear-clip, and a sensor arm. The headset's reference and ground electrodes are on the ear clip and the EEG electrode is on the sensor arm, resting on the forehead above the eye. It safely measures and outputs the EEG power spectrums (Alpha waves, Beta waves, Theta waves, Delta waves), NeuroSkyeSense meters (attention and mediation) and eye blinks. The headset is rated at 60 Hz. Neurosky device has been selected for our experiments because of its affordable price and its light weight that enables the user to wear it for a long time although it has only one delectrode which limits the features carried in the data acquired.

B. Dataset collection

Data were collected from many subjects. During the course of work, it was proven that training with seven subjects was enough for the objective. Accordingly, brain signals of seven subjects have been recorded in a quiet room to prevent any possible distractions. A reading session consists of 14 readings for one subject, each reading 14 seconds recording of EEG signals. The subject closes his eyes to prevent eye-blink artifacts. Moreover, the subject should not move any muscle to prevent muscle artifacts. The readings are further used to train different classifiers.

C. Recording Setup

Subjects were all Egyptians whose mother tongue is Arabic. They were not under any medication and had no diseases. Their ages were between 22 and 23. Each subject was presented to 14 “Yes” questions and 14 “No” questions.

The subject sits on a chair in a quiet room and wears the headset that is connected with the laptop via Bluetooth.

The subject is asked a Y/N question. He presses the “start button”.

The subject starts to think of the answer, while the recording is processing for 14 seconds.

The Neurosky support development kit (SDK) drops the data recorded on the first two seconds and the last two seconds to ensure subject’s concentration.

III. Feature extraction

The NeuroSky SDK offers the values of low Alpha, high Alpha, low Beta, high Beta, Delta, Gamma and Theta waves. It gives a single value for each wave every one second. Alpha and Beta waves are more related to mental activities and thinking. Low alpha, high alpha, low beta and high beta were selected to be the classifying features. Delta waves were discarded because they only appear in baby brain waves, Gamma waves were discarded because they appear during recognizing object or sound tasks, and Theta also were discarded, as they only appear in young children or during idle tasks. The minimum, maximum and average values for each of the 4 considered waves were calculated giving a feature vector of 12 values for each sample. Ten samples were acquired.

The NeuroSky SDK also offers the value of Raw EEG data acquired, with a sampling rate of 512 Hz. Based on the work of Ting et al. [8], wavelet packet decomposition was applied to get 6-level decomposition tree using the Sym8 wavelet, and the first 6 nodes from level 6, which have the respective frequency bands (0-8, 8-16, 16-24, 24-32, 32-40, 40-48) Hz were selected. The average coefficient and the band energy for each node were calculated which gives a feature vector of 12 values for each sample. The next phase is to input the vectors to the classifiers.

IV. Classification

Four classifiers and ensemble network were used:

1) Support Vector Machine (SVM)

Matlab function svmtrain was chosen with a variety of kernel functions: linear, Quadratic, RBF with sigma values 0.2, 0.4, 0.6, 0.8, 1 and Polynomial. Matlab code, e.g.,

```
struct = svmtrain(train , target_train , 'kernel', 'rbf', 'rbf_sigma', 0.2);
output = svmclassify(struct , test2);
```

With SVM, as the number of classes increases, the prediction time increases significantly. Also, with large classes, the training time increases significantly.

2) Discriminant Analysis (DA)

It was shown by Shashua [9], that the decision hyper planes for binary classification obtained by SVMs is equivalent to the solution obtained by Fisher’s linear discriminant on the set of support vectors. It was also shown by Gallinari et al. [10], that the neural networks classifiers are equivalent to DA. That justifies using DA algorithm for classification of our work. Both linear and quadratic functions were used for classification.

3) Self-Organizing Map (SOM)

An enhancement step was added to the SOM architecture: First, data are classified into the maximum number of clusters (unsupervised). Second, these clusters are mapped in a supervised mode into one of 2 clusters based on majority concept.

4) Feed Forward Back-propagation (FFBP)

Various multi layer networks were used, each with different layer-neuron combinations. This was done by trial and error starting with single layer-5 neuron network ending with 156 networks combinations.

5) Ensemble network

Ensemble networks or combining multiple classifiers aim to reduce generalization error and to improve the classification performance over individual classifiers, as has been presented by Avnimelech and Intrator [11], Hansen and Salamon [12], Hashem and Schmeiser [13], and Sharkey [14]. Two types of combinational networks were tested. Each used the same classifiers as above: DA, SOM and SVM. The 2 types are: 1- Two-stage network with the combinations SOM→DA, SOM→SOM, SOM→SVM. It means that the output of SOM is fed as input for DA, SOM and SVM networks, respectively 2- Voting network composed of SOM, DA and SVM, where simple majority voting was applied on the outputs of the classifiers.

V. Testing and results

Data collected are placed in 12 files. Each file’s data are arranged in one of two ways: Firstly, Random arrangement, where “Yes” and “No” are scattered in a random way in the file. This type of files will be named “randomly arranged file”. Secondly, Sequence arrangement, where “Yes” is placed before “No” or vice versa. This type of files will be named “sequentially arranged file’. Testing was carried out in 2 modes: 1- Offline mode: separating data acquisition and classification phases into two groups. One group is for

 ISBN: 978-1-61208-340-7

classification and the other group is used for testing. Signals used in offline were primarily concerned with the 7 subjects and the conditions mentioned in the earlier section. 2- Online mode: based on prior training of the network, new data signals from various subjects are examined for classification. Signals used covered different subjects to validate the work done. Subjects participating in testing were of different genders and ages. Experiment conditions also varied, where it involved quiet, as well as noisy environment, and eyes open also. Classifiers were applied to raw data, as well as Alpha and Beta signals.

A. *Offline testing with Raw Data*

The following are the hit rates achieved when using raw data with different classifiers. The charts show the result for sample tested files, while the Tables show the min, max and average values. The testing was carried with 60%-40% train-test ratios based on different trials' results.

1) *DA classifier*

DA classifier is not affected by the data arrangement, so, results from random files are the same as sequential files. Figure 1 shows the average hit rates for 6 files with various random data arrangements using linear and quadratic functions. For two files, both functions gave equal results. Linear function was better in 3 files, while quadratic was better in one. So, it could be concluded that linear function gave on average better results. Table I shows the min, max and average values for DA classifier.

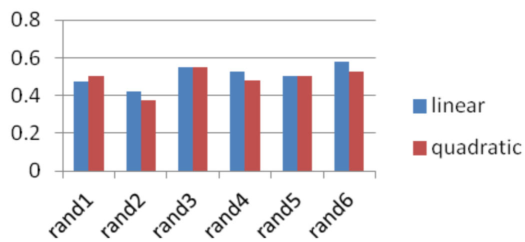

Figure 1. Hit rates for DA classifier

TABLE I. MIN, MAX AND AVERAGE VALUES FOR DA CLASSIFIER

Min value	Max value	Average value
0.3717	0.5769	0.49843

2) *SOM classifier*

A modification was done on regular SOM network as follows: 1- samples are first classified into the largest possible number of clusters, n. Starting with n= 2, then increasing the number of clusters until classification remains the same in two successive iterations. 2- the n clusters are remapped to two clusters with majority rule, i.e., classes with "Yes" samples greater than "No" are mapped to "Yes" cluster and the same for "No". A fragment code is shown in figure 2.

```
// train phase.  Constructor with rate 0.7
        Som = new SOMCSharp(0.7)
// preparing the input
        Som.prepare_C(train, test);
// start is training main method. It takes learning rate and continues to repeat training until stable conditions.
// we start with level=1 i.e., 2 clusters
// check for stability conditions applied, if "Yes stop training
// else increase level (number of clusters)
        Som.start(0.7);
// start method calls leveltrain() method which starts the weights with a fixed value of 0.5 and update weights in train() method according to distances
   // test phase where test() method computes hit rate
   int SOMout = Som.test_C(data);
```

Figure 2. Fragment code showing flow of instructions

Figure 3 shows the average hit rates for 6 files, with various data arrangements and various sigma, using SOM classifier. The tests show that for sigma= 0.5, 0.7, 0.9, the best results are obtained, except for one file. Table II shows the min, max and average values achieved.

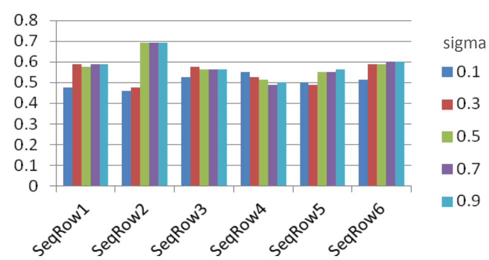

Figure 3. Hit rates for SOM classifer.

TABLE II. MIN, MAX AND AVERAGE VALUES FOR SOM CLASSIFIER

Min	Max	Average
0.4615	0.6923	0.55854

3) *SVM classifier*

Matlab was used to test and train networks. SVM classifier experienced no changes when the arrangement of data changed from random to sequential. Figure 4 shows the values with different mapping functions and permutations (P1-P6).

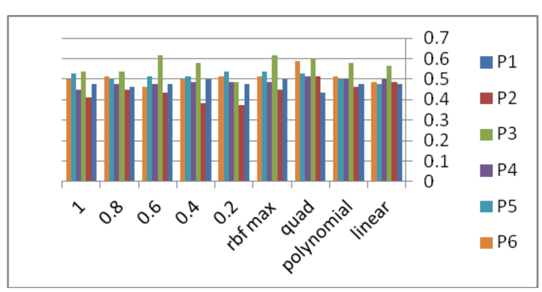

Figure 4. Hit rates for SVM classifier.

Copyright (c) IARIA, 2014. ISBN: 978-1-61208-340-7

Table III shows the min, max and average values for SVM classifier.

TABLE III. MIN, MAX AND AVERAGE VALUES FOR SVM CLASSIFIER

Min value	Max value	Average value
0.3718	0.6154	0.4962

4) *FFBP*

Different networks have been tried starting from 1 layer/5 neurons, 1 layer/10 neurons, 2 layers/5 neurons, 2 layers/10 neurons, 3 layers/ 5 neuron and 10 neurons reaching 156 combinations. Also various training functions like trainbr, trainbfg, traincgb were applied. In all the combinations, the best result achieved was way far from the other networks. It was concluded that FFBP network results should not be listed.

B. *Offline testing with Alpha and Beta signals*

The same percentage of train-test data was used; 60%-40% with the same classifiers but using the Alpha and Beta signals. Results are as follows:

TABLE IV. OFFLINE RESULTS WITH ALPHA AND BETA SIGNLS

Network	Min value	Max value	Average value
DA	0.5147	0.641	0.448
SOM	0.4744	0.6795	0.5599
SVM	0.4231	0.641	0.5531

Comparing results in Table IV and in Tables I-III, we conclude that the best average offline results were obtained from SOM networks with Raw data and with Alpha and Beta signals. It proves that the modification done on the network enhanced the performance with offline testing. Average efficiencies were 55.8 and 55.9%. Maximum values were 69 and 67.9%.

Ensemble networks were also tested and the results are presented below.

1) *Ensemble Networks*

Alpha and Beta signals were used in offline testing with Ensemble networks. In the ensemble multistage network, the first network was always SOM based on the results shown in Table IV. Slightly better average result was achieved, 0.5666.

For the voting system that comprised of DA network, SVM network, and SOM network and based on majority function, average result further improved to reach 0.5933. The improvement in both cases was on the expense of time. Results are shown in Table V.

Accordingly, for offline recognition of unspoken two words, the most suitable network is a Voting system with simple majority function applied on Alpha and Beta signals. The Voting network comprises DA, SVM, and SOM.

TABLE V. ENSEMBLE NETWORKS' RESULTS

Network	Min value	Max value	Average value
SOM → DA	0.42	0.5512	0.4902
SOM →SOM	0.6538	0.4487	0.5516
SOM →SVM	0.4261	0.6391	0.5666
VOTING system	0.551	0.6667	0.5933

C. *Online testing*

Online testing was carried out using the DA, SVM, SOM and Voting networks on male and female subjects of age groups 19-23 years to identify the network that gives the best result. The testing environment was a college hall with open windows and open door having 40-60 students. Table VI shows the average efficiencies for this testing.

TABLE VI. AVERAGE ONLINE EFFICIENCIES

Network	Efficiency
DA	0.485
SVM	0.60
SOM	0.49
VOTING	0.51

It could be seen that DA, SOM and Voting gave nearly equal results, while SVM gave a remarkably better result. Further online testing in the same environment was carried out using SVM to find average result. Table VII shows the results.

TABLE VII. ONLINE TESTING WITH SVM ON MALE/FEMALE SUBJECTS

Male/Female	No of Subjects	Min value	Max value	Average value
Male	17	0.3	0.9	0.564
Female	10	0.3	0.8	0.57

The average efficiency is less than that attained by offline. This is an expected result, although the maximum value reached in online testing is much higher.

1) *Random subjects*

The online system was further tested in a two day exhibition (Egyptian Engineering Day) with 60-80 subjects per day. Subjects were males and females visitors between 19-23 years old. They stopped randomly to try the system. The recording environment was the exhibition booth, while asking the people around to stay quiet, as there was enough noise from the surroundings. The result was 56.7%.

 ISBN: 978-1-61208-340-7

VI. Validation

Our target was to help disabled young children to communicate. Children need basically "Yes" and "No" words to interact. The device used in this work is light weight with only 1 electrode. The electrode is put directly on the forehead. Signal transfer is done via Bluetooth. It is very suitable for disabled adults also, as in case mentioned by Graimann et al. [1]. All previous work listed hereunder was based on 5 words recognition using a multi electrode device. The cap used for recordings is made of spandex type fabric and is equipped with many electrodes, up to 128. The electrodes have to be filled with conductive gel. The subject wears the cap and it has to be tight. The cap is then attached to the subject with straps that are connected to a band, which is attached around the supper part of the body. The target subject for those researches was normal adults like astronauts, divers, soldiers in battle who need to communicate. Difference in target subjects justifies the difference in number of words to be recognized. Table VIII represents a summary of results for the work presented and other work.

TABLE VIII. Summary of the current work with previous work

Research	Scope	Average efficiency
Wester and Shultz [2]	5 words 21subjects 16 electrodes	- 42%
Calliess [3]	5 words 16 electrodes used 23 subjects	- 15.5 % better than Wester and Shultz [2] (49%)
Porbadnigk [4]	5 words 16 electrodes used 23 subjects	- 45.5% for block - By chance for separate words
Porbadnigk et al. [5]	5 words 21subjects 16 electrodes	- 45.5% for block - By chance for separate words
Torres-García et al. [6]	5 words 21 subjects 16 electrodes used	- 20%
Our work	2 words 7 subjects 1 electrode used	- Offline 56% - Online 57%

VII. Conclusion and Future Work

In this work, different ways of recognizing unspoken speech have been investigated using a single electrode EEG device. The unspoken speech comprised of 2 words; "Yes" and "No". Seven subjects were used for training. The hit rate for unspoken speech recognition depended on the subject's concentration and absence of artifacts. A modification to SOM network classification was made by making the classification a two step process. This improved the results of the SOM network in offline testing. Offline average hit rate of 59% was reached. An online average hit rate of 57% was achieved. It is worth mentioning that the ensemble network performed well only in offline, while failed in online, which is contrary to Avnimelech and Intrator [11], Hansen and Salamon [12], Hashem and Schmeiser [13], and Sharkey [14]. Other researches in the same field used multi-electrode EEG devices (up to 16 electrodes) to recognize 5 words with average recognition rates ranging between 20% and 49%. The future work is to use a single electrode device to recognize more than 2 words since such a device is much easier and lighter to wear and to adjust for certain cases as the disabled children.

References

[1] B. Graimann, B. Allison, and G. Pfurtscheller, "Brain–computer interfaces: a gentle introduction", The Frontiers Collection, pp. 1-27, Springer-Verlag Berlin Heidelberg, 2010, doi :10.1007/978-3-642-02091-9_1.

[2] M. Wester and T. Schultz, "Unspoken speech -speech recognition based on electroencephalography". Masters thesis, Institut für Theoretische Informatik Universität Karlsruhe (TH), Karlsruhe, Germany, 2006.

[3] J. Calliess, "Further investigations on unspoken speech. - findings in an attempt of developing EEG-based word recognition", Bachelor work, Interactive Systems Laboratories Carnegie Mellon University, Pittsburgh, PA, USA and Institut fuer Theoretische Informatik Universitaet Karlsruhe (TH), Karlsruhe, Germany, 2006.

[4] A. Porbadnigk, "Eeg-based speech recognition: impact of experimental design on performance". Institut fuer Algorithmen und Kognitive Systeme, Universitaet Karlsruhe (TH), Karlsruhe, Germany, 2008.

[5] A. Porbadnigk, M. Wester, J. Calliess, and T. Schultz, "EEG-based speech recognition: impact of temporal effects", 2nd International Conference on Bio-inspired Systems and Siganl Procesing (Biosignals 2009), Porto, Portugal, pp. 376-381.

[6] A. Torres-García, C.A. Reyes-García, and L. Pineda, "Toward a silent speech interface based on unspoken speech," Proc. Biosignals 2012 (BIOSTEC), Vilamoura, Algarve, Portugal, SciTePress, Feb 2012, pp. 370-373.

[7] http://neurosky.com/products-markets/eeg-biosensors/hardware/, April 2014.

[8] W. Ting, Y. Guo-zheng, Y. Bang-hua, and S. Hong, "EEG feature extraction based on wavelet packet decomposition for brain computer interface", ScienceDirect, Measurement 41, (2008), pp. 618-625. Available online from: www.sciencedirect.com.

[9] A. Shashua, "On the equivalence between the support vector machine for classification and sparsified Fisher's linear discriminant". Neural Process Lett 9(2):129–139, 1999.

[10] P. Gallinari et al., "On the relations between discriminant analysis and multilayer perceptrons". Journal Neural Networks, vol. 4 issue 3, pp. 349–360, 1991, doi: 10.1016/0893-6080(91)90071-C.

[11] R. Avnimelech and N. Intrator, "Boosted mixture of experts: an ensemble learning scheme," Neural Computation., vol. 11, pp. 483–497, Feb 1999, MIT press, doi: 10.1162/089976699300016737.

[12] L. Hansen and P. Salamon, "Neural network ensembles," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 12, no. 10, pp. 993–1001, Oct 1990, B.001.

[13] S. Hashem and Schmeiser, "Improving model accuracy using optimal linear combinations of trained neural networks," IEEE Trans. Neural Networks, vol. 6, no. 3, pp. 792–794, May 1995, doi: 10.1109/72.377990.

[14] A. Sharkey, "Multi-net systems, combining artificial neural nets: ensemble and modular multi-net systems", Berlin, Germany: Springer-Verlag, 1999, pp. 1–30.