REPORT DOCUMENTATION PAGE

Form Approved OMB NO. 0704-0188

1. REPORT DATE (DD-MM-YYYY)	2. REPORT TYPE	3. DATES COVERED (From - To)
21-07-2010	Final Report	1-Oct-2009 - 30-Jun-2010

4. TITLE AND SUBTITLE

Advanced Signal Processing and Machine Learning Approaches for EEG Analysis

5a. CONTRACT NUMBER
W911NF-09-1-0491

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
611102

6. AUTHORS

Vijayakumar Bhagavatula

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES AND ADDRESSES

Carnegie Mellon University
Office of Sponsored Programs
Carnegie Mellon University
Pittsburgh, PA 15213 -

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Research Office
P.O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSOR/MONITOR'S ACRONYM(S)
ARO

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
57221-CS-II.1

12. DISTRIBUTION AVAILIBILITY STATEMENT

Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

The views, opinions and/or findings contained in this report are those of the author(s) and should not contrued as an official Department of the Army position, policy or decision, unless so designated by other documentation.

14. ABSTRACT

Electroencephalography (EEG) offers a non-invasive brain-imaging technology with potential to extract user intent from brain signals. This can offer a potential method for dispersed soldiers to communicate silently with one another. The usual interface for acquiring EEG signals may house 128 or more electrodes. Each EEG signal may be sampled at KHz sampling rates and may last for a few seconds. Thus the number of samples used to represent each trial can be large. The goal of this short-term innovative research (STIR) project was to investigate innovative

15. SUBJECT TERMS

EEG signals, machine learning, signal processing

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	15. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Vijayakumar Bhagavatula
UU	UU	UU	UU		19b. TELEPHONE NUMBER 412-268-3026

Report Title

Advanced Signal Processing and Machine Learning Approaches for EEG Analysis

ABSTRACT

Electroencephalography (EEG) offers a non-invasive brain-imaging technology with potential to extract user intent from brain signals. This can offer a potential method for dispersed soldiers to communicate silently with one another. The usual interface for acquiring EEG signals may house 128 or more electrodes. Each EEG signal may be sampled at KHz sampling rates and may last for a few seconds. Thus the number of samples used to represent each trial can be large. The goal of this short-term innovative research (STIR) project was to investigate innovative sample and channel (i.e., EEG electrode) selection methods to reduce the storage and computational complexity in analyzing EEG signals. In experiments aimed at determining the redundancy in imagined speech EEG signals, it was observed that EEG data has limited spatial redundancy, but large temporal redundancy. In another set of experiments, we investigated the classification of two imagined speech syllables (namely “Ba” and “Ku”) from imagined speech EEG signals. Using all good channels, only one of the seven volunteer subjects produced "better than chance" classification accuracy of about 60%. By selecting specific electrodes, two subjects yielded better-than-chance results with recognition rates close to 60% for all trials. Overall classification rates appear to have improved slightly by the selection of electrodes, indicating that imagined speech classification performance can be improved by careful selection of EEG electrodes.

List of papers submitted or published that acknowledge ARO support during this reporting period. List the papers, including journal references, in the following categories:

(a) Papers published in peer-reviewed journals (N/A for none)

Number of Papers published in peer-reviewed journals: 0.00

(b) Papers published in non-peer-reviewed journals or in conference proceedings (N/A for none)

Number of Papers published in non peer-reviewed journals: 0.00

(c) Presentations

Number of Presentations: 0.00

Non Peer-Reviewed Conference Proceeding publications (other than abstracts):

Number of Non Peer-Reviewed Conference Proceeding publications (other than abstracts): 0

Peer-Reviewed Conference Proceeding publications (other than abstracts):

Number of Peer-Reviewed Conference Proceeding publications (other than abstracts): 0

(d) Manuscripts

Number of Manuscripts: 0.00

Patents Submitted

Patents Awarded

Graduate Students

NAME	PERCENT_SUPPORTED
Euiseok Hwang	0.40
Andres Rodriguez	0.20
FTE Equivalent:	**0.60**
Total Number:	**2**

Names of Post Doctorates

NAME	PERCENT_SUPPORTED
FTE Equivalent:	
Total Number:	

Names of Faculty Supported

NAME	PERCENT_SUPPORTED	National Academy Member
Vijayakumar Bhagavatula	0.11	No
FTE Equivalent:	**0.11**	
Total Number:	**1**	

Names of Under Graduate students supported

NAME	PERCENT_SUPPORTED
FTE Equivalent:	
Total Number:	

Student Metrics

This section only applies to graduating undergraduates supported by this agreement in this reporting period

The number of undergraduates funded by this agreement who graduated during this period:	0.00
The number of undergraduates funded by this agreement who graduated during this period with a degree in science, mathematics, engineering, or technology fields:.....	0.00
The number of undergraduates funded by your agreement who graduated during this period and will continue to pursue a graduate or Ph.D. degree in science, mathematics, engineering, or technology fields:.....	0.00
Number of graduating undergraduates who achieved a 3.5 GPA to 4.0 (4.0 max scale):	0.00
Number of graduating undergraduates funded by a DoD funded Center of Excellence grant for Education, Research and Engineering:	0.00
The number of undergraduates funded by your agreement who graduated during this period and intend to work for the Department of Defense	0.00
The number of undergraduates funded by your agreement who graduated during this period and will receive scholarships or fellowships for further studies in science, mathematics, engineering or technology fields:	0.00

Names of Personnel receiving masters degrees

NAME

Total Number:

Names of personnel receiving PHDs

NAME

Total Number:

Names of other research staff

NAME PERCENT_SUPPORTED

FTE Equivalent:

Total Number:

Sub Contractors (DD882)

Inventions (DD882)

Final Project Report (FPR) for ARO Project W911NF0910491

Advanced Signal Processing and Machine Learning Approaches for EEG Analysis

Prof. Vijayakumar Bhagavatula,
Department of Electrical and Computer Engineering,
Carnegie Mellon University,
Pittsburgh, PA 15213

Period of performance: 10/01/2009 – 06/30/2010

Table of Contents

List of Figures

List of Tables

1 STATEMENT OF THE PROBLEM

Electroencephalography (EEG) offers a non-invasive brain-imaging technology with potential to extract user intent from brain signals. This can offer a potential method for dispersed soldiers to communicate silently with one another. Army-supported MURI (led by University of California, Irvine) on "*Silent Spatialized Communication among Dispersed Forces*" is aimed at developing this technology.

One interface for acquiring EEG signals is an EEG skull cap that may house 128 or more electrodes. Each EEG signal may be sampled at KHz sampling rates and may last for a few seconds. Thus the number of samples used to represent each trial can be in the order of millions. Given the multiple trials, multiple subjects and multiple types of experiments necessary for developing effective classification techniques, the number of overall samples can become very large leading to significant computational and storage complexity challenges. Even worse, this may represent the case where much of the data (corresponding to electrodes placed in some regions) may be irrelevant and even nuisance signals for the covert speech classification problem at hand. Thus, the goal of this short-term innovative research (STIR) project was to investigate innovative sample and channel (i.e., EEG electrode) selection methods to reduce the storage and computational complexity in analyzing EEG signals.

2 SUMMARY OF THE MOST IMPORTANT RESULTS

First set of experiments were aimed at determining the redundancy in imagined speech EEG signals. This was done through the application of compressed sensing and sparse representation concepts as well as through manual selection. We observed that EEG data has limited spatial redundancy, e.g., while the number of electrodes is 110, we seem to need about 75 non-zero coefficients, implying that the spatial redundancy is less than 50%. EEG data appears to have large temporal redundancy. An EEG signal set with 880 samples is well represented by about 167 non-zero coefficients, corresponding to about 80% temporal redundancy.

In the second set of experiments, we investigated the classification of imagined speech syllables "Ba" and "Ku" from imagined speech EEG signals collected from seven subjects at University of California, Irvine. Using all "good" channels, almost all of the subjects with the exception of subject 7 produce chance results. Subject 7 produced 2-class classification accuracy of about 60%. We also investigated manual electrode selection and automatic electrode selection. To select electrodes using the automated method, the electrode correlations are first computed using earlier trials as templates. Then for each cluster found by the automated method, a single electrode from the cluster is selected as the main electrode, which is the one located in closest to the center of the cluster. From the group of main electrodes, the final selected electrodes are then limited to ones that lie above the brain regions that are activated during speech production. The classification rates from the manual and automatic channel selection are comparable, although the automated method selects fewer electrodes. Subjects 2 and 7 yielded better-

than-chance results with recognition rates close to 60% for all trials, and subjects 3 and 6 were slightly better than chance. Overall classification rates appear to have improved slightly by selecting specific electrodes.

3 DATA COLLECTION

3.1 Imagined Speech EEG Data

The imagined speech EEG dataset was collected in the Department of Cognitive Sciences at UCI. They conducted experiments in which volunteer subjects imagined speaking two syllables, /ba/ or /ku/ while their electrical brainwave activity was being recorded by EEG. These syllables were selected since they contain no semantic meaning so that classification would be performed on the imagined speech instead of the semantic contribution to imagined speech production [1]. The subjects were instructed to covertly speak a given syllable at a certain rhythm, both of which were provided via audio cues. So in each trial, a syllable (either /ba/ or /ku/) was heard through a set of Stax electrostatic earphones followed by a series of clicks at the desired rhythm for the imagined speech. Approximately 1.5 seconds after the last click, the subject was to begin to imagine speaking the spoken syllable at the given rhythm (see Figure 1 for a timeline [1]). During the time segment corresponding to EEG signals of interest, no audio or video stimuli were present - the subject was supposed to imagine speaking that syllable at that rhythm.

As described in [1], the EEG data were recorded using a 128 Channel Sensor Net by Electrical Geodesics [2] and sampled at 1024Hz. A single experimental session was typically comprised of 20 trials for each condition, and data were recorded over separate sessions, which varied for each subject. During the recording, the subjects were seated in a dimly lit room and instructed to keep their eyes open and to fixate on a certain point while avoiding any eye blinks and muscle movement.

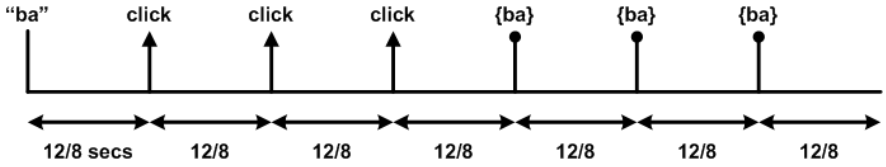


Figure 1. Timeline for a single trial in the covert speech experiment [1]

3.2 Multi-class Motor Imagery EEG Data

Multi-class motor imagery EEG data from the BCI Competition III (dataset IIIa) [3] was also used to test the general applicability of the proposed channel selection method to EEG data for imagined tasks. The subject was cued to either imagine left hand, right hand, foot, or tongue movements for a total of 4 different classes of data. A 64-channel EEG amplifier from Neuroscan was used to record brainwave activity, and the EEG was sampled at 250Hz and filtered to a frequency range of 1 to 50 Hz [3]. Sixty EEG channels were recorded, and indexed according to Figure 2.

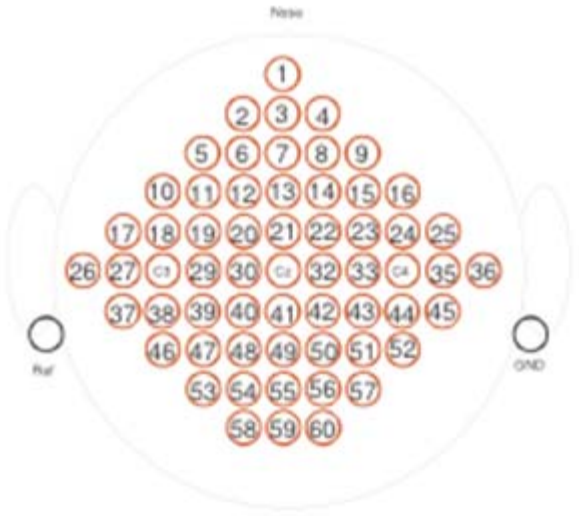


Figure 2. Position of EEG electrodes for the multi-class motor imagery EEG dataset [3].

4 DATA PREPROCESSING

During these experiments, although the subjects attempt to keep movement to a minimum during these recordings, the EEG data inevitably contain some presence of artifacts (i.e., changes in EEG amplitudes that do not correspond to brainwave activity but eye movements or muscle movements instead). These artifacts tend to dominate and obscure the actual cortical signal. Additionally, in some cases these artifacts can be fairly predictive. This may result in deceptively high recognition rates since a classifier would succeed by identifying these artifacts as opposed to the portions of the signal that reflect the true brainwave activity. Therefore, the EEG data is first preprocessed to remove artifacts and also to reduce noise (e.g., 60Hz line noise).

For the imagined speech EEG data, electromyographic (EMG) artifacts (i.e., muscle artifacts) are first considered for removal using the same preprocessing steps suggested by D'Zmura et al. in [1]. EEG signals from 18 of the 128 electrodes that are closest to the neck, eyes, and temple are discarded since they are the most prone to EMG artifacts. Furthermore, since EMG artifacts are typically present in frequencies greater than 25Hz, the remaining EEG signals are filtered to a frequency range of 4 to 25Hz, which additionally removes the 60Hz line noise from these signals. The data is then detrended to remove baseline drift and downsampled to a more manageable sampling rate of 256Hz. In addition, 4 electrodes were found to be faulty in a number of trials (as no data were collected by these electrodes during these trials), so signals from these electrodes were completely discarded as well.

For the imagined movement EEG data, artifact information was provided by the group that collected the data, so trials containing artifacts were already visually identified and flagged. This given information was used to discard contaminated trials.

5 COMPRESSED SENSING FOR IMAGINED SPEECH EEG DATA

A major research breakthrough in the past five years has been the concept of compressed sensing [4]. It has been shown that sparse signals (i.e., signals which have a small number of non-small values in some domain) can be accurately represented using a small number of projections of such signals on to data-independent random vectors. The signal of interest does not have to be sparse in the original signal domain --- it may be sparse in some other domain such as the frequency domain or in discrete cosine transform (DCT) domain. The reconstruction from such a sparse representation can be achieved using L1 optimization methods. For the EEG-based covert speech classification task, we have investigated the benefits, if any, of compressed sensing.

5.1 Simulation Setup

As EEG signals have both spatial redundancy and temporal redundancy, they are processed as a 2-D image. As shown in Figure 3, every signal set has K time samples and the data dimension is $N_{channel} \times K$, where $N_{channel}$ is the number of channels.

In the Ba-Ku imagined speech syllable classification experiments, signals from Subject 6 led to the best classification results. Hence we focused on data from Subject 6 to investigate the benefits of compressed sensing theory. In the following simulations, both the training set and the testing set are from Subject 6 for the imagined speech EEG data.

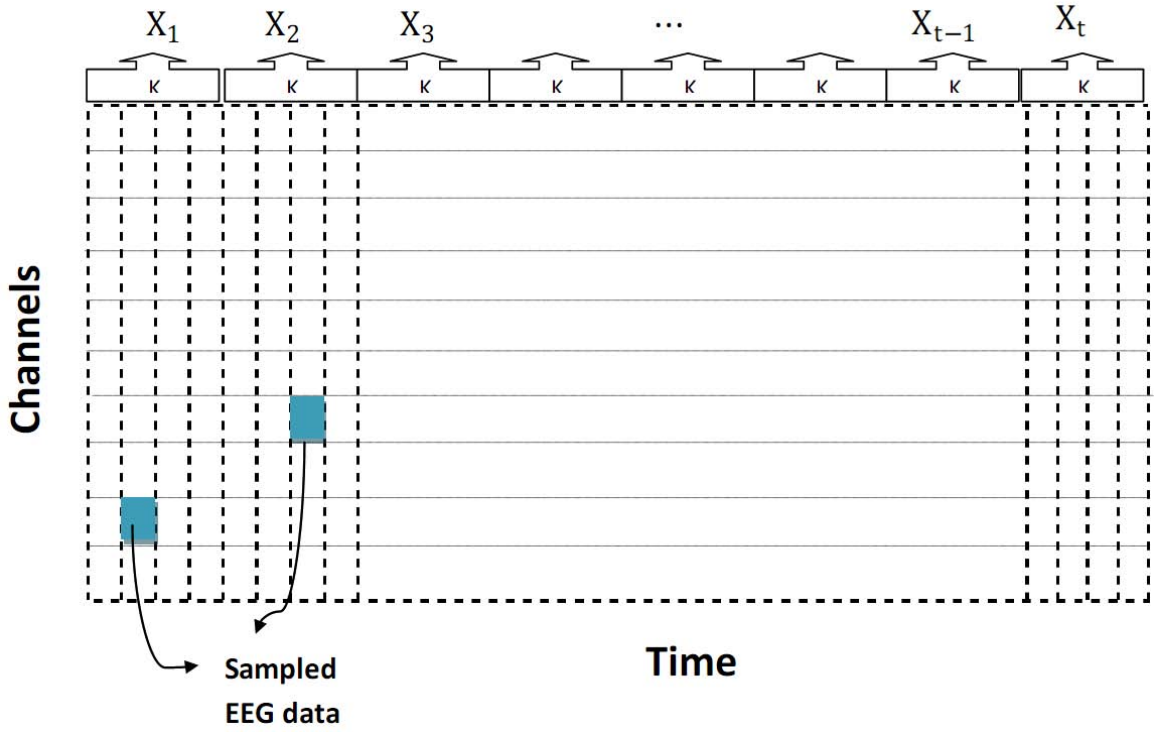


Figure 3. Segmentation of original EEG data to small sets suitable for sparse analysis. The EEG signals are treated as a 2-D spatial-temporal image. The dimension for each set is $N_{channel} \times K$, where $N_{channel}$ is the number of channels and K is number of time samples.

5.2 Sparse Representation for EEG signals

In sparse representation, signals are described as linear combinations of a few "atoms" from a pre-specified dictionary. The key problem here is how to find the suitable dictionary. Classically, such dictionaries are built from a data model, such as the Discrete Cosine Transform (DCT) or the Discrete Wavelet Transform (DWT) for natural images. As EEG signals tend to be quite noisy, a good general model is hard to obtain. So we chose to learn a dictionary from the training set [5]-[7]. The corresponding optimization problem is:

$$\begin{aligned} &\min_{B,S} \frac{1}{2\sigma^2}\|X - BS\|_F^2 + \beta\sum_{i,j}\phi(S_{i,j}) \\ &s.t: \sum_i B_{i,j}^2 \le c, \forall j = 1,...,n \end{aligned} \tag{1}$$

where X is the $m \times t$ input matrix (each column is an input vector), B is the $m \times n$ dictionary (each column is a basis vector) and S is the $n \times t$ coefficient matrix. The penalty function $\varphi(\cdot)$ is L_1 or epsilon L_1 norms, defined as follows.

$$\phi(S_{i,j}) = \begin{cases} |S_{i,j}|_1 & (L_1 \ \ penalty) \\ \left(S_{i,j}^2 + \varepsilon_s\right)^{1/2} & (Epsilon \ \ L_1 \ \ penalty) \end{cases} \tag{2}$$

In (1), the first part corresponds to the representation ability of the dictionary B, and the second penalty part denotes the sparsity of the representation. The algorithm in [6] is adopted to solve this optimization problem. The key idea in the algorithm is - though the original problem (1) is non-convex, it is convex in B with S fixed, and convex in S with B fixed. So an iterative approach is used by solving the two convex sub-problems alternately.

5.2.1 Results

10,000 samples are used to train the dictionary, and another 200 samples are used for testing. Here the input vector is taken as shown in Figure 3, with $K \in \{1, 3, 5, 8\}$. The simulation result is shown in Figure 4. Following observations can be made from these simulation results.

- EEG data has limited spatial redundancy. When $K = 1$ (i.e., when there are only spatial samples), the average number of non-zero coefficients is about 75 and number of electrodes is 110. More than half of the coefficients are non-zero.
- EEG data has large temporal redundancy. The compression ratio increases with the number of time samples, K. When signal size becomes 880 ($K = 8$), it is perfectly represented by about 167 non-zero coefficients.

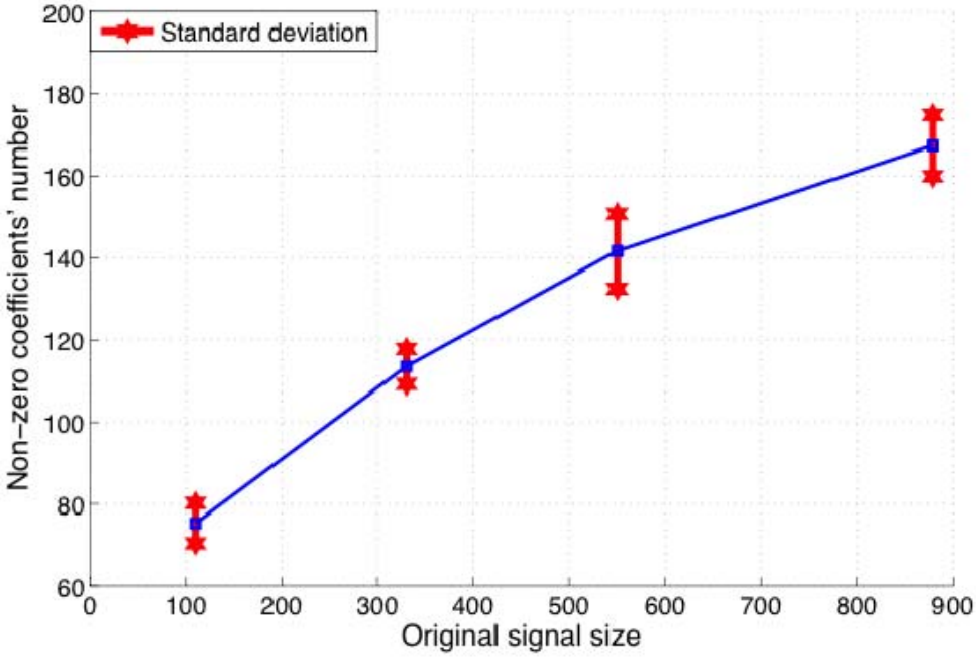


Figure 4. Sparse representation on optimal dictionary. Here the x-axis is the dimension of original signal, the y-axis is the number of non-zero coefficients from decomposing the signal with the dictionary. The result is an average, and the red lines denote the standard deviation on test data.

5.3 Compressed Sensing

In this section the classical compressed sensing method presented in [8] is investigated. Consider a general linear measurement process that computes the inner product between original signal x and a collection of vectors $\{\Phi_j\}$:

$$y = \Phi x = \Phi B s \qquad (3)$$

where y is the measurement data, Φ is defined as the measurement matrix (which in practice is drawn at random), B is the dictionary, and s is the sparse vector. Basis pursuit is used to get the sparse vector s, which is formulated as:

$$\hat{s} = \arg\min_{s} \left\{ \left\| y - \Phi s \right\|_2 + \lambda \left\| s \right\|_1 \right\} \qquad (4)$$

This optimization problem is solved with the method described in [9].

The simulation setup is the same with last section. Define $\hat{x}$ as the reconstructed signal. To analyze the result quantitatively, the reconstruction error defined as:

$$\varepsilon = \frac{\left\| \hat{x} - x \right\|_2}{\left\| x \right\|_2} \qquad (5)$$

which serves as the performance index. The simulation results are shown in Table 1. The EEG data are well reconstructed by compressed sensing. When K = 8, the EEG data can be compressed by a ratio of 5.5.

	$\frac{m}{r}$	$E(\varepsilon)$	$STD(\varepsilon)$
m = 110, r = 90	1.22	3.74%	0.0249
m = 110, r = 80	1.38	4.84%	0.0342
m = 330, r = 120	2.75	4.33%	0.0151
m = 330, r = 110	3.00	5.13%	0.0195
m = 330, r = 100	3.30	7.11%	0.0250
m = 550, r = 150	3.67	3.87%	0.0679
m = 550, r = 140	3.93	4.77%	0.0809
m = 550, r = 130	4.23	5.76%	0.0826
m = 880, r = 180	4.89	3.92%	0.0201
m = 880, r = 170	4.89	4.63%	0.0212
m = 880, r = 160	4.89	5.44%	0.0254

Table 1. Simulation results for compressed sensing.

Here, m is the EEG signal size, r is the measurement data size, m/r is the compression ratio, $E(.)$ is the expectation, and $STD(.)$ is the standard deviation.

5.4 Channel Selection

The purpose of channel selection is to reconstruct the brainwave signal with as few electrodes as possible. Different from classical compressed sensing discussed in last subsection, the measurement matrix Φ here is restricted to be a "selection" matrix; the elements of the measurement matrix are either zero or one, and the sum of each row is one. A direct approach for getting optimal Φ is to randomly pick channels and select the channel set with minimum error. However, such a "brute force" method is very slow. As an alternative, an iterative algorithm is proposed. The cost function is designed to minimize the error of reconstruction over training data. The key idea is that given an initial channel set, we apply compressed sensing and add the worst channel into the set iteratively. The detailed algorithm is shown in Table 2. The simulation result over 1,000 test samples is shown in Figure 5. By selecting 80 out of 110 channels, we observe a 5% reconstruction error whereas if we allow the number of channels to increase to 90, this reconstruction error decreased to 3%.

	Input: initial channel set Θ, training data $\{X_i\}$, $i \in \{1,\ldots, t\}$, target channel number r
1	Make measurement matrix Φ from channel sets Θ.
2	Use compressed sensing to get the reconstruction $\{\hat{X}_i\}$, $i \in \{1,..., t\}$
3	For channel j, calculate the error vector $v_j = \frac{\left\|\hat{X}_{\cdot,j} - X_{\cdot,j}\right\|_2}{\left\|X_{\cdot,j}\right\|_2}$, $j \in \{1,..., m\}$
4	Add $p = \arg\max_{j \notin \Theta} v_j$ into Θ
5	If size(Θ) < r, go back to step 1; else stop.

Table 2. Channel selection algorithm

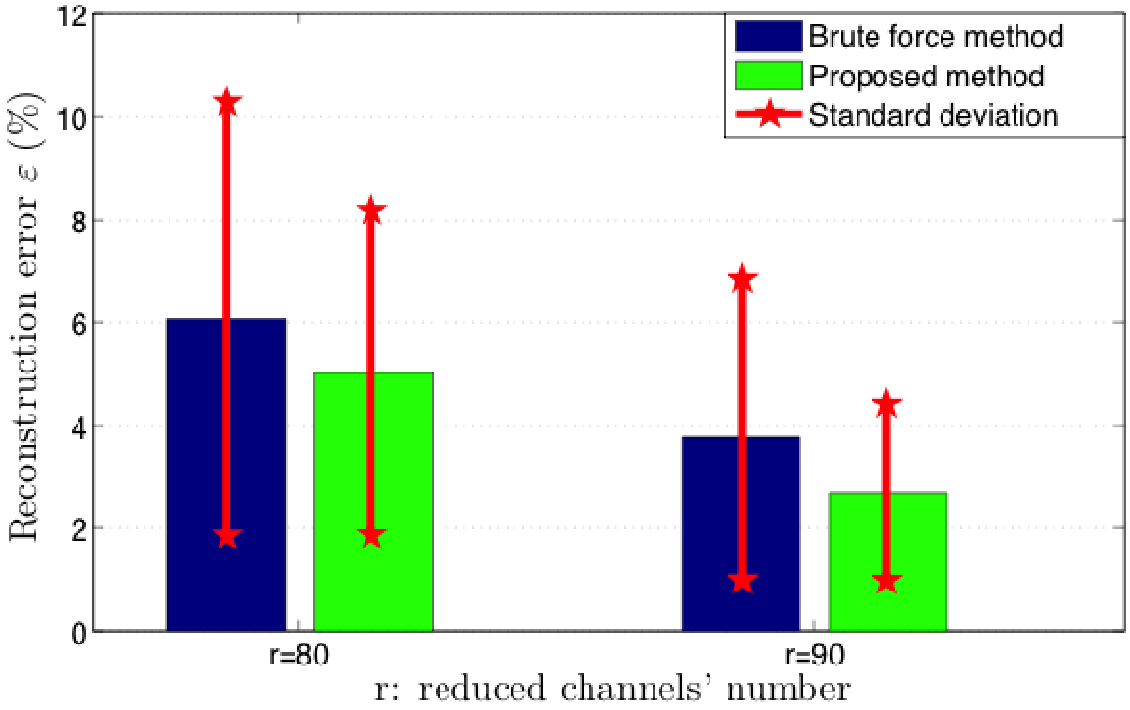


Figure 5. Simulation result for channel selection. *r* is final number of selected channels (out of 110), and ε is the reconstruction error

6 CHANNEL SELECTION FOR EEG SIGNAL CLASSIFICATION

In this section, we will describe other approaches to reducing the number of channels (i.e., EEG electrodes). Channel reduction may also be helpful for classification as well, by either reducing the amount of computation required by discarding electrodes, or by enhancing the signal classification by only using channels thought to contain information relevant to the signal of interest. The approach for classifying imagined speech EEG data is first presented along with the classification results for using all electrodes, and then classification results will also be shown where specific electrodes are selected either manually or automatically. Lastly, it will be shown that the automatic electrode selection approach can also be applied to discard redundant electrodes, as will be demonstrated using motor imagery EEG data.

6.1 Imagined Speech EEG Classification

6.1.1 Preprocessing

For the imagined speech EEG data, preprocessing is performed as described in Section 4, but the sub-sampling rate is increased from 5 to 16, and the "bad" electrodes that are closest to the eyes, neck and temple are used (instead of being removed from further consideration) to denoise the remaining EEG signals from the "good" electrodes.

6.1.2 Denoising

Define the brain signal S as the electrical signal from brain activity, and the noise א as the signal that is contaminated with artifacts. Two problems are discussed in this section: is

there is noise in data, and if so, how can we remove it. Before going into detail however, some basic assumptions are necessary:

1. Noise א is statistically independent of the brain signal S.
2. The "bad" channels only contain noise.
3. Both the noise and the brain signals are Gaussian distributed and the observation system is linear.

The first assumption is reasonable as artifacts are thought to originate from independent biological processes, and are therefore independent of the brain activity of interest. The second assumption holds because the "bad" channels are far from the active brain region and heavily contaminated by artifacts. The third assumption is a simplification of the complicated reality.
Based on these three assumptions, the system is modeled as:

$$X_{bad} = A_{bad}\aleph \quad (6)$$

$$X_{good} = BS + A_{good}\aleph \quad (7)$$

Eq. (6) comes from assumption 2, where X_{bad} represents the signals from the "bad" channels, and A_{bad} is the corresponding mixing matrix. In Eq. (7), X_{good} represents the signals from the "good" channels, B is brain signal's mixing matrix, and A_{good} is mixing matrix for noise from the "good" channels. The noise א and the brain signal S are assumed to satisfy the following equations:

$$E(\aleph\aleph^{\mathrm{T}}) = I \quad (8)$$

$$E(\aleph S^{\mathrm{T}}) = 0 \quad (9)$$

where E(·) denotes the statistical expectation operator, and I is the identity matrix of appropriate size.

6.1.3 Existence of noise in EEG data

If a "good" channel is correlated with some "bad" channel, then it is likely to be contaminated. For example, if we look at channel 88, Figure 6 shows the correlation coefficients between the "good" channel 88 and the 14 bad channels, where the correlation coefficient between two signals Y and Z, $\rho_{Y,Z}$, is defined as follows:

$$\rho_{Y,Z} = \frac{E[(Y - E(Y))(Z - E(Z))]}{\sigma_Y \sigma_Z} \quad (10)$$

where σ_Y and σ_Z is the standard deviation of Y and Z respectively.

The maximum absolute correlation coefficient in this example is 0.53, which indicates there is a strong presence of noise in channel 88. The result is similar for most other "good" channels. Therefore, if we want to use these channels for classification, then noise reduction is necessary.

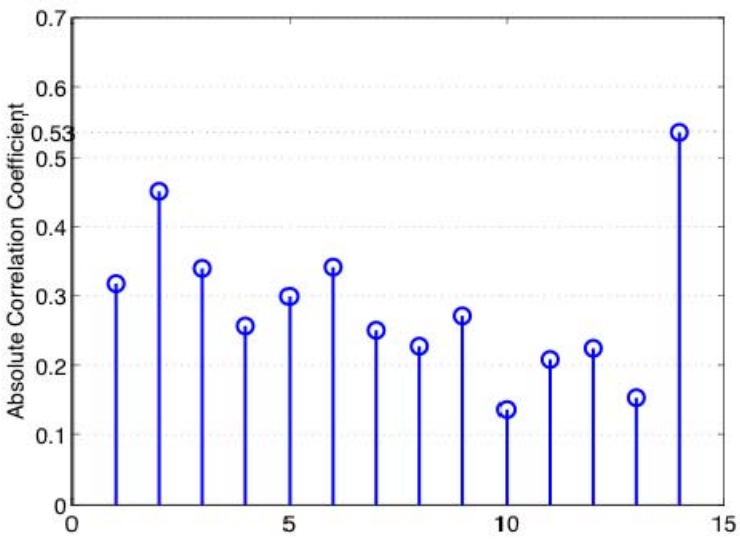


Figure 6. Absolute correlation coefficients between "good" channel 88 and 14 "bad" channels.

6.1.4 Subspace Denoise Method

This subspace method decomposes the noisy EEG signals into noise and uncontaminated brain signals. Figure 7 provides an illustration of this decomposition. Here, the noise space is defined as the subspace spanned by noise, and the signal space is defined as the subspace spanned by brain signals. Based on the independence assumption, the signal space is orthogonal to the noise space. So the decomposition may be geometrically interpreted as projection on different subspace.

Let $U_M \in R^{MxN}$ denote an orthogonal basis of the noise space, where M is the space dimension and N is number of samples. The projection of a noisy signal X onto the noise space is defined as:

$$X_n = U_M^{\mathrm{T}} U_M X \qquad (11)$$

Here X_n can be taken as the noisy part of contaminated signal X. The original brain signal is then defined as:

$$X_{denoised} = X - X_n = (I - U_M^{\mathrm{T}} U_M) X \qquad (12)$$

In implementation, the noise space is derived from singular value decomposition (SVD) of the "bad" channels, i.e.,

$$[U, D, V] = SVD(X_{bad}) \qquad (13)$$

The first M rows of U provide an orthogonal basis of the noise space. As M increases, more noise is removed. However, the linear assumption here is an approximation of the real system. A large value of M risks the possibility of losing useful information. Figure 8 shows how the denoising result varies with M. In the current implementation, M = 4.

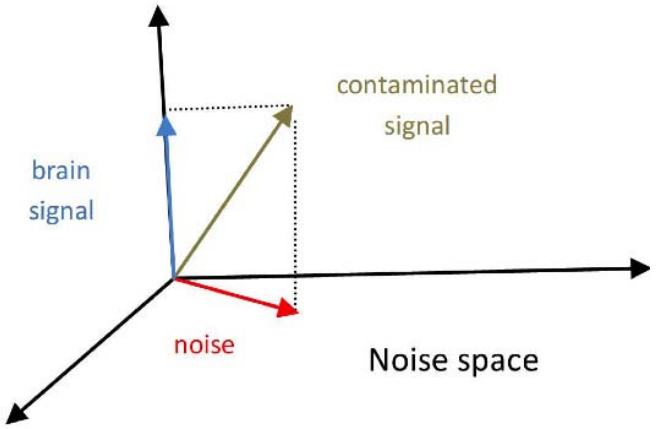


Figure 7. An illustration of subspace denoise method

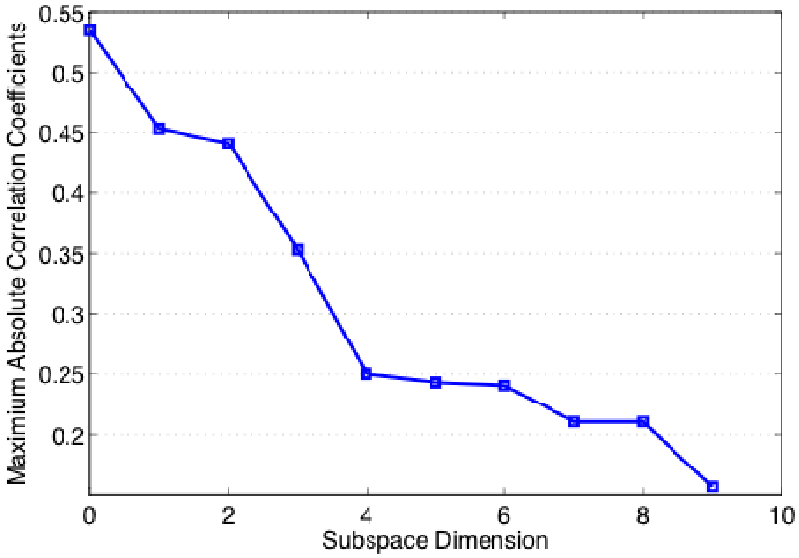


Figure 8. Absolute correlation coefficients between "good" channel 88 and 14 "bad" channels after applying the subspace denoise method. The x-axis denotes the subspace dimension *M*, and the y-axis is the maximum absolute correlation coefficient of channel 88 with the bad channels.

6.1.5 Feature Extraction

There are still two open questions after preprocessing and denoising. First, there is no clock tick to synchronize when subject is covertly speaking. So the expected time stamp of each syllable is not accurate. Such an error would accumulate for the second and third syllables in the same trial, which makes the estimated times for when the syllable is covertly spoken unreliable. To compensate for this, only a 0.2 second time window around the first syllable is kept in each trial, as shown in Figure 9. Furthermore, the feature dimension is still too large. For example, there are 1404 points left in each trial but the number of trials of subject 2 is only 116. This may result in overfitting the classifier. So before classification, Principal Component Analysis (PCA) is used to reduce the feature dimension from 1404 to 2.

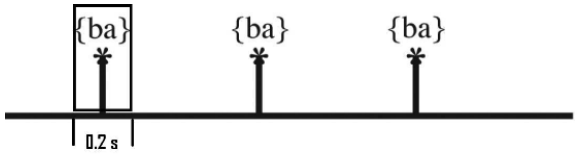


Figure 9. Only the data in 0.2s time window around the first syllable of each trial is retained to build the feature set.

6.1.6 Classification

A Support Vector Machine (SVM) with a quadratic kernel function is used here as the final classifier. The experimental results are shown in Table 3, where all of the classification rates are averaged over 20 iterations of 5-fold cross-validation, where the training and testing set are kept separate.

6.1.7 Results

Using all "good" channels, almost all of the subjects with the exception of subject 7 produce chance results.

Subject	Dataset Size	Training Accuracy	Testing Accuracy
S1	ba: 119, ku: 118	0.5548	0.5042
S2	ba: 116, ku: 116	0.5248	0.5041
S3	ba: 200, ku: 203	0.5245	0.4855
S4	ba: 187, ku: 189	0.5365	0.4930
S6	ba: 79, ku: 79	0.5199	0.4856
S7	ba: 80, ku: 79	0.6128	0.6000

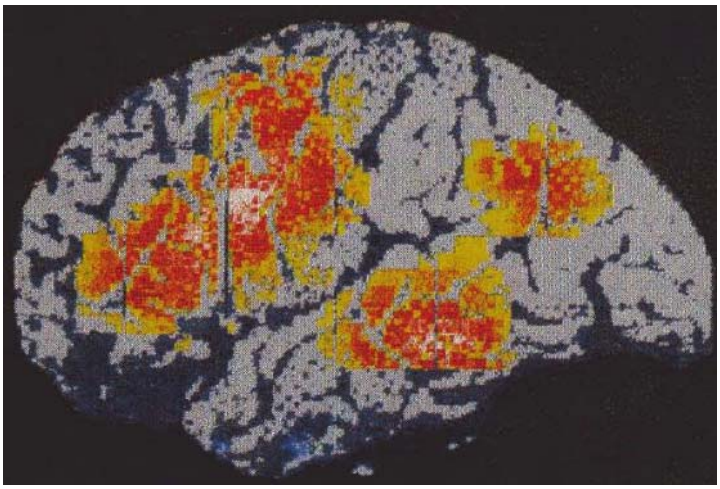
Table 3. Classification results for 6 subjects using all "good" channels

6.2 Channel Selection

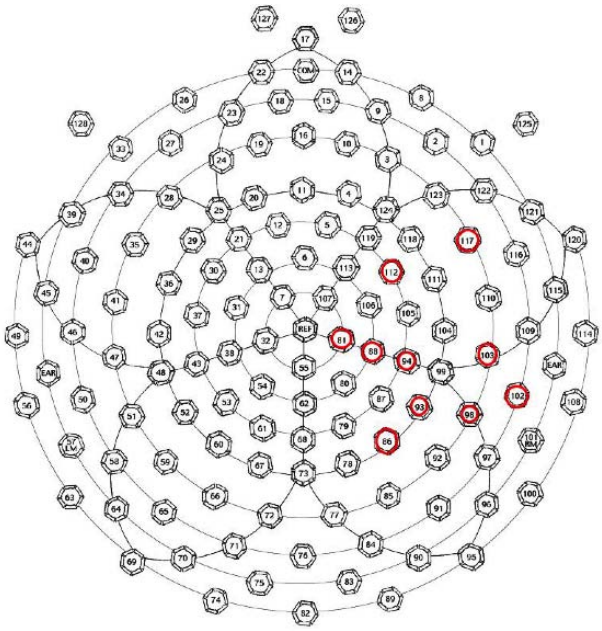
6.2.1 Manual Channel Selection for Imagined Speech

Neuroscience research has shown that different brain regions control different human behaviors. Speaking covertly is believed to activate the frontal cortex as well as Broca's and Wernicke's areas [10], as shown in Figure 10(a). The electrodes that are distant from the active region, such as those directly at the top or the back of head, may not provide any relevant information. Discarding these electrodes would furthermore considerably reduce the number of electrodes.

However, exact coordinates of active regions are not provided. Therefore, we select 10 electrodes roughly near the possible active region, as shown in Figure 10(b). Experimental results show that even such an imprecise set up can still achieve reasonable results for certain subjects.



(a)



(b)

Figure 10. (a) Specific regions of the cortex involved in covert speech, identified with PET scanning. (b) EEG electrode distribution. The ten selected electrodes are marked by red circles.

6.2.2 Automatic Channel Selection for Imagined Speech

Alternatively, electrodes may be automatically selected based on the information provided in their signals. EEG is known to have poor spatial resolution, so adjacent electrodes tend to be highly correlated. The channels may therefore be clustered based on the correlation between electrodes, and this clustering may also reveal location information about where stronger signals may be found.

In this approach, correlation coefficients are calculated for each electrode pair by first normalizing the denoised signals, $X_{denoised}$, to unit variance, and then computing the

covariance matrix. If the correlation coefficient between two electrodes is greater than some threshold α, then the two electrodes may potentially belong to the same cluster. A list of potential clusters can then be constructed for each electrode based on how highly correlated each electrode is with the others. That is, by thresholding the covariance matrix C, we now have a list of N clusters where N is the number of channels, and electrodes n and m belongs to the same cluster if $C(n, m) > \alpha$.

However, since it is possible to have electrodes that are not highly correlated with each other to belong the same cluster, where they may both instead be correlated with another electrode, a co-occurrence matrix is subsequently built based on the initial clustering.

The co-occurrence matrix C_M is built as follows:

$$C_M(n, m) = \sum_{i=n,\, i\neq m}^{N} \left(\mathrm{I}\left(C(n,i) > \alpha \;\; \& \;\; C(i,m) > \alpha\right)\right), \text{ for } n, m = 1, ..., N \qquad (14)$$

where N is the number of channels, C is the covariance matrix of the normalized denoised signals $X_{denoised}$, α is the correlation threshold (which is set to 0.6 for the imagined speech EEG data), and I(·) is the indicator function. Each element in this co-occurrence matrix indicates how many times electrode n co-occurs with electrode m in the initial set of clusters. Each row then (or column, since C_M is symmetric) is a potential cluster, and contains a list of which electrodes belong to it, where the index of the nonzero elements in that row or column denotes the electrode that is in that cluster. C_M now represents a more complete set of potential clusters than simply using $C > \alpha$.

However, this cluster list still needs to be pared down since we still have N clusters, and there are clearly going to be highly similar clusters. We can agglomerate similar clusters by using the symmetry of the co-occurrence matrix to help find related clusters. For example, if the set A contains the indices of all electrodes belonging to potential cluster 1, then potential cluster 1 is related to the corresponding set of clusters whose indices match those contained in the set A. That is, for a given electrode n we have:

$$A = \underset{i \in [1, N]}{\arg\max}\ \mathrm{I}(C_M(n, i) > 0) \qquad (15)$$

where I(·) is again the indicator function such that:

$$\mathrm{I}(C_M(n, i) > 0) = \begin{cases} 1, & \text{Electrode } n \text{ co-occurs with Electrode } i \\ 0, & \text{Electrode } n \text{ does not co-occur with Electrode } i \end{cases} \qquad (16)$$

and A is the set of clusters to which electrode n belongs. Therefore, we will consider all clusters in the set A to be "related" in that they may be similar clusters. We will further expand this set by also considering other electrodes found in the cluster set A, and form a new set, B:

$$B = \underset{j \in [1, N]}{\arg\max}\ \mathrm{I}(C_M(j, i) > 0) \text{ for } \forall i \in A \qquad (17)$$

Here, B is the set of all clusters that relate to the electrodes of the clusters found in set A.

So, for example, let us consider a set of initial clusters:

Cluster 1 = {Electrodes 1, 2, 5, 6}
Cluster 2 = {Electrodes 1, 2, 5, 6, 7}
Cluster 3 = {Electrodes 3, 4}
Cluster 4 = {Electrodes 3, 4, 6}
Cluster 5 = {Electrodes 1, 2, 5, 6, 7}
Cluster 6 = {Electrodes 1, 2, 4, 5, 6, 7}
Cluster 7 = {Electrodes 2, 5, 6, 7}

We will start by selecting a cluster, say Cluster 1. We would then derive the set *A* of related clusters, as Clusters 1, 2, 5, and 6. Then, to expand the set *A* to obtain the set *B*, we consider all unique electrodes in Clusters 1, 2, 5, and 6:

Cluster 1 = {Electrodes 1, 2, 5, 6}
Cluster 2 = {Electrodes 1, 2, 5, 6, 7}
Cluster 5 = {Electrodes 1, 2, 5, 6, 7}
Cluster 6 = {Electrodes 1, 2, 4, 5, 6, 7}

namely Electrodes 1, 2, 4, 5, 6, and 7. Therefore, since Electrodes 4 and 7 are included in this set, we should also consider Clusters 4 and 7, so we add that to set *A* to grow to our new set $B = \{1, 2, 4, 5, 6, 7\}$.

Now we will define the probability of an electrode *n* belonging the overall cluster that is most representative of *B* as:

$$P(e_n \in \text{Final Cluster}) = \frac{1}{b_n} \sum_{k \in B} \frac{|A_k|}{|A|} \left(\mathrm{I}(C_M(n, k) > 0) \right) \qquad (18)$$

where e_n is electrode *n*, b_n is the number of times electrode *n* appears within the cluster set *B*, $|A_k|$ is the number of times electrode *k* appears in set *A*, $|A|$ is the number of elements in the set *B*, and the function $\mathrm{I}(C_M(n, k) > 0)$ represents whether or not electrode *n* is in cluster *k*.

A threshold is then used to determine if an electrode belongs to the final cluster derived from set *B*. In this study, the clustering threshold is set to 0.8. Once the electrodes are finally declared to belong to a cluster, the related clusters to the final cluster will no longer be used, and the remaining clusters will be analyzed to form a new *B* set. And if none of the electrodes have probabilities that lie above this clustering threshold, then no cluster is found for this set, and the next cluster is selected to form a new set *B*. This process continues until all clusters have either been discarded or used. In the given example, the final clusters found are:

Cluster 1 = {Electrodes 1, 2, 5, 6, 7}
Cluster 2 = {Electrodes 3, 4}

6.2.3 Results from Channel Selection over All Trials

Results from the manual electrode selection and the automatic electrode selection method discussed in the previous section are presented here. To select electrodes using the automated method, the electrode correlations are first computed using earlier trials as templates (e.g., Trials 1 and 21 from each class). Then for each cluster found by the automated method, a single electrode from the cluster is selected as the main electrode, which is the one located in closest to the center of the cluster. From the group of main electrodes, the final selected electrodes are then limited to ones that lie above the brain regions mentioned in Section 6.2.1, and this electrode region is shown in Figure 11. In manually or automatically selecting channels that are thought to lie above areas of the brain that are activated during speech production, the algorithm described in Section 6.1 is able to achieve better-than-chance results. The classification rates from the manual and automatic channel selection are comparable, although the automated method selects fewer electrodes. The classification results are summarized in Table 4 and Table 5, and plotted in Figure 12.

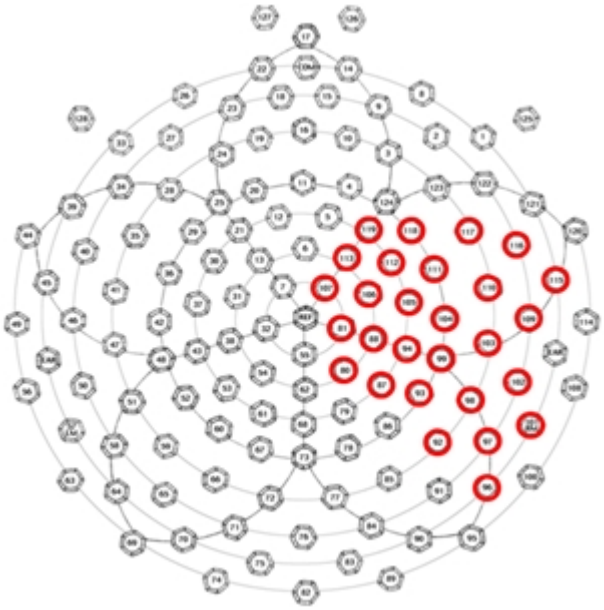


Figure 11. Final channel selection is limited only to the electrodes marked by the red circles in automatic electrode selection.

Manual Selection

Subject	Dataset Size	Training Accuracy	Testing Accuracy
S1	ba: 119, ku: 118	0.5661	0.5371
S2	ba: 116, ku: 116	0.6240	0.5952
S3	ba: 99, ku: 98	0.5602	0.5029
S4	ba: 118, ku: 119	0.5409	0.4829
S6	ba: 79, ku: 79	0.6124	0.5697
S7	ba: 80, ku: 79	0.6129	0.6045

Table 4. Classification results for 6 subjects using manually selected channels

Automatic Selection				
Subject	**Dataset Size**	**Training Accuracy**	**Testing Accuracy**	**# Channels**
S1	ba: 119, ku: 118	0.5625	0.5296	8
S2	ba: 116, ku: 116	0.6310	0.5938	7
S3	ba: 99, ku: 98	0.5940	0.5667	3
S4	ba: 118, ku: 119	0.5577	0.5216	8
S6	ba: 79, ku: 79	0.5935	0.5500	5
S7	ba: 80, ku: 79	0.6125	0.6090	1

Table 5. Classification results for 6 subjects using automatically selected channels

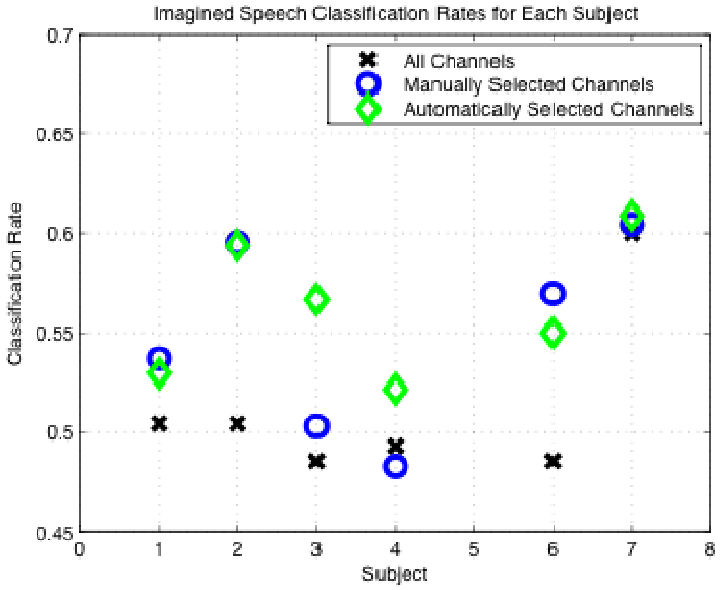


Figure 12. Imagined Speech Classification Rate for each subject, using all channels, manually selected channels, or automatically selected channels

Subjects 2 and 7 yielded better-than-chance results with rates of close to 60% for all trials, and subjects 3 and 6 were slightly over chance. Overall classification rates have improved by selecting specific electrodes. Classification was also performed for each session, for each subject. These results are shown in Table 6 to Table 17.

		Manual Selection		
Subject	**Session**	**Dataset Size**	**Training Accuracy**	**Testing Accuracy**
S1	1	ba: 20, ku: 20	0.5860	0.3305
	2	ba: 20, ku: 18	0.5853	0.4146
	3	ba: 19, ku: 20	0.7004	0.5555
	4	ba: 20, ku: 20	0.6712	0.4880
	5	ba: 20, ku: 20	0.6724	0.5464
	6	ba: 20, ku: 20	0.6298	0.4310

Table 6. Classification results for each session for Subject 1 using manually selected channels

		Automatic Selection			
Subject	**Session**	**Dataset Size**	**Training Accuracy**	**Testing Accuracy**	**# Channels Selected**
S1	1	ba: 20, ku: 20	0.5905	0.3896	3
	2	ba: 20, ku: 18	0.6170	0.4297	2
	3	ba: 19, ku: 20	0.7063	0.5860	4
	4	ba: 20, ku: 20	0.6617	0.5384	4
	5	ba: 20, ku: 20	0.6884	0.5487	2
	6	ba: 20, ku: 20	0.6494	0.4838	2

Table 7. Classification results for each session for Subject 1 using automatically selected channels

		Manual Selection		
Subject	**Session**	**Dataset Size**	**Training Accuracy**	**Testing Accuracy**
S2	1	ba: 19, ku: 20	0.6758	0.5100
	2	ba: 20, ku: 20	0.5825	0.3893
	3	ba: 19, ku: 19	0.7020	0.5580
	4	ba: 20, ku: 20	0.6326	0.4597
	5	ba: 20, ku: 19	0.7323	0.6071
	6	ba: 18, ku: 18	0.6385	0.4540

Table 8. Classification results for each session for Subject 2 using manually selected channels

		Automatic Selection			
Subject	**Session**	**Dataset Size**	**Training Accuracy**	**Testing Accuracy**	**# Channels Selected**
S2	1	ba: 19, ku: 20	0.7672	0.6559	3
	2	ba: 20, ku: 20	0.6002	0.4387	3
	3	ba: 19, ku: 19	0.6106	0.4754	4
	4	ba: 20, ku: 20	0.6739	0.5119	6
	5	ba: 20, ku: 19	0.7215	0.5844	7
	6	ba: 18, ku: 18	0.6949	0.4995	7

Table 9. Classification results for each session for Subject 2 using automatically selected channels

		Manual Selection		
Subject	**Session**	**Dataset Size**	**Training Accuracy**	**Testing Accuracy**
S3	1	ba: 19, ku: 20	0.6480	0.5279
	2	ba: 20, ku: 20	0.7138	0.5712
	3	ba: 19, ku: 19	0.6324	0.4761
	4	ba: 20, ku: 20	0.6379	0.5080
	5	ba: 20, ku: 19	0.6688	0.5024

Table 10. Classification results for each session for Subject 3 using manually selected channels

		Automatic Selection			
Subject	**Session**	**Dataset Size**	**Training Accuracy**	**Testing Accuracy**	**# Channels Selected**
S3	1	ba: 19, ku: 20	0.6160	0.4686	4
	2	ba: 20, ku: 20	0.6863	0.5905	1
	3	ba: 19, ku: 19	0.6458	0.4836	1
	4	ba: 20, ku: 20	0.7593	0.6549	2
	5	ba: 20, ku: 19	0.6820	0.5563	2

Table 11. Classification results for each session for Subject 3 using automatically selected channels

		Manual Selection		
Subject	**Session**	**Dataset Size**	**Training Accuracy**	**Testing Accuracy**
S4	1	ba: 20, ku: 20	0.6619	0.5608
	2	ba: 18, ku: 20	0.6897	0.5115
	3	ba: 20, ku: 20	0.7288	0.6038
	4	ba: 20, ku: 20	0.6123	0.4256
	5	ba: 20, ku: 19	0.5945	0.4487
	6	ba: 20, ku: 20	0.5927	0.4002

Table 12. Classification results for each session for Subject 4 using manually selected channels

		Automatic Selection			
Subject	**Session**	**Dataset Size**	**Training Accuracy**	**Testing Accuracy**	**# Channels Selected**
S4	1	ba: 20, ku: 20	0.6232	0.4435	3
	2	ba: 18, ku: 20	0.6987	0.5826	2
	3	ba: 20, ku: 20	0.6478	0.5039	3
	4	ba: 20, ku: 20	0.6334	0.4897	2
	5	ba: 20, ku: 19	0.	0.	--
	6	ba: 20, ku: 20	0.	0.	--

Table 13. Classification results for each session for Subject 4 using automatically selected channels. Sessions 5 and 6 are blank as no electrodes were selected for these sessions.

		Manual Selection		
Subject	**Session**	**Dataset Size**	**Training Accuracy**	**Testing Accuracy**
S6	1	ba: 20, ku: 20	0.6760	0.5536
	2	ba: 20, ku: 19	0.5850	0.3811
	3	ba: 19, ku: 20	0.6386	0.4853
	4	ba: 20, ku: 20	0.5941	0.4145

Table 14. Classification results for each session for Subject 6 using manually selected channels

		Automatic Selection			
Subject	**Session**	**Dataset Size**	**Training Accuracy**	**Testing Accuracy**	**# Channels Selected**
S6	1	ba: 20, ku: 20	0.6772	0.5261	5
	2	ba: 20, ku: 19	0.5925	0.4338	5
	3	ba: 19, ku: 20	0.6776	0.5856	6
	4	ba: 20, ku: 20	0.7160	0.6226	5

Table 15. Classification results for each session for Subject 6 using automatically selected channels

		Manual Selection		
Subject	**Session**	**Dataset Size**	**Training Accuracy**	**Testing Accuracy**
S7	1	ba: 20, ku: 20	0.6916	0.5330
	2	ba: 20, ku: 18	0.9005	0.8470
	3	ba: 19, ku: 20	0.6028	0.4568
	4	ba: 20, ku: 20	0.7400	0.6151

Table 16. Classification results for each session for Subject 7 using manually selected channels

		Automatic Selection			
Subject	**Session**	**Dataset Size**	**Training Accuracy**	**Testing Accuracy**	**# Channels Selected**
S7	1	ba: 20, ku: 20	0.7379	0.6488	1
	2	ba: 20, ku: 18	0.9042	0.8344	1
	3	ba: 19, ku: 20	0.7045	0.5861	1
	4	ba: 20, ku: 20	0.7263	0.5842	1

Table 17. Classification results for each session for Subject 7 using automatically selected channels

Overall, the classification rates varied from session to session for each subject. Some sessions lead to better results than others, most notably for subject 7. Subject 7 seemed to have the best results per session and overall. It was also interesting to note that the automatic channel selection method selected one channel for the in-session and overall classification, which also happened to be the channel with the best classification rate if

the classification were to be performed for each individual channel. Classification rates were calculated for each channel and were plotted for each electrode (shown in Figure 13). The red 'X' in the plot marks the classification rate for the electrode that was selected for each session and overall. This seems to indicate that useful information may indeed lie in the electrode regions suggested in this study, and that correlation information may be used to locate interesting areas of brain activity.

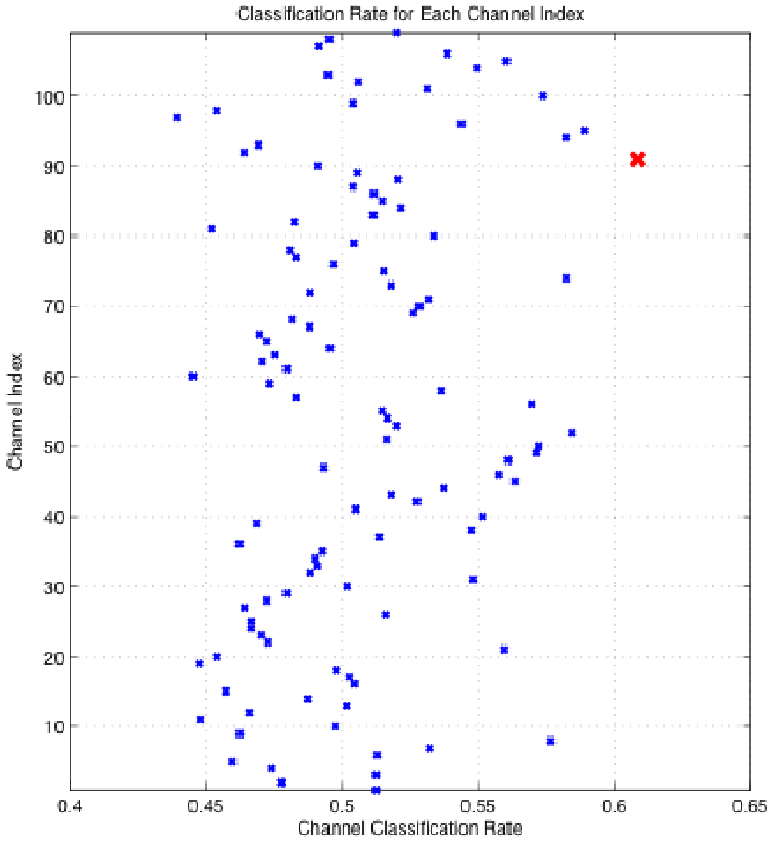


Figure 13. Classification rates per channel for Subject 7

6.3 Imagined Movement EEG Classification

The automated channel selection method may also be used to discard redundant electrodes, or electrodes that do not seem to provide information that may be helpful for classification. The EEG data for motor imagery was used to test the ability of the automated channel selection method to find a reduced set of electrodes that would either maintain or improve upon the classification rates using all electrodes.

6.3.1 Feature Extraction and Classification

For the motor imagery EEG data, we used the signal model shown in Figure 14 along with some assumptions to estimate the power spectral density (PSD) of each EEG signal (here represented by $x[n]$, the observed signal).

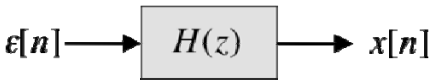


Figure 14. Block diagram of the signal model

With this model, we assume that the EEG signals can be modeled as a wide-sense stationary random process and that each signal is generated by inputting a zero-mean white noise $\varepsilon[n]$ with variance σ^2 into a linear shift-invariant all-pole filter. The corresponding time series model [11] is given in equation (19).

$$x[n] = -\sum_{k=1}^{p} \alpha_k x[n-k] + \varepsilon[n] \qquad (19)$$

where $x[n]$ is the observed signal at time n and α_k are the model coefficients. The integer p is the order of this model. As can be seen in this equation, this autoregressive (AR) model attempts to predict the current time sample given previous time samples, and its transfer function is as given as follows.

$$H(z) = \frac{1}{1 + \sum_{k=1}^{p} \alpha_k z^{-k}} \qquad (20)$$

Consequently, the AR coefficients α_k completely determine the spectrum of the model output, since the spectrum of the model output is the product $|H(e^{j\omega})|^2$ and σ^2. The AR coefficients thus characterize the spectral peaks of the signal and its sharpness.

AR coefficients were computed for each electrode's signal using the Burg method as described in [12] and concatenated to form a feature vector. Orders 2 through 6 were tested to see which order gave the best classification accuracies. An AR model order of 3 appeared to be optimal for the imagined movement EEG dataset.

The imagined movements were then classified using a SVM with a polynomial kernel of degree 6, with a "one against the rest" scheme to classify the 4 different classes. The publicly available software LibSVM [13] was used for SVM classification.

6.3.2 Results

In using all 60 electrodes for subject k3b with 360 trials, 20 iterations of 5-fold cross validation were run, resulting in an average classification rate of 82.06%. The automated channel selection method selected the 29 electrodes marked in green shown in Figure 15, and only using these electrodes yielded an average classification accuracy of 81.88%, which is comparable to the results using all electrodes. This demonstrates the ability of the algorithm to find electrodes that contain information useful for classification.

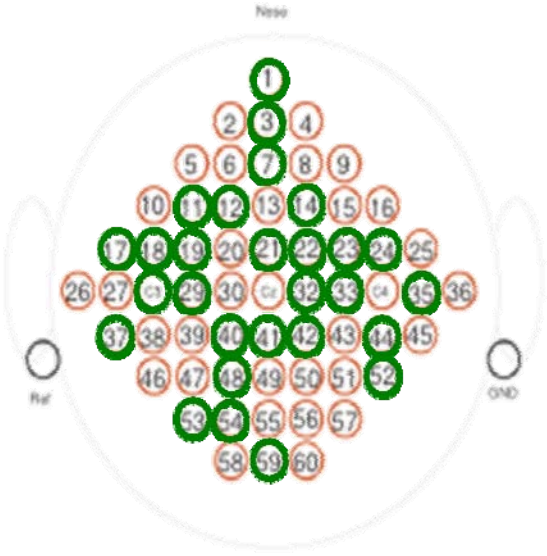


Figure 15. Topography of electrodes for the motor imagery EEG data. Electrodes selected by the automated channel selection method are marked in green.

7 CONCLUSIONS

In this report we investigated the effects of reducing the number of electrodes and the number of samples per electrode in EEG signal-based classification. First we discussed the algorithm for building suitable dictionary, which can sparsely represent EEG data. Then the capability of compressed sensing to reconstruct whole EEG signal is tested from reduced measurement data. We also studied the feasibility of using geometric information and results from neuroscience knowledge to reduce the number of channels. It was shown that manually or automatically selecting electrodes based on neuroscience knowledge yields better classification results. After subspace denoising, the selected electrodes achieve near or above 60% classification accuracy on almost half the subjects. However, because of the lack of exact location information and limited number of subjects, there is need for more research.

8 BIBLIOGRAPHY

[1] M. D'Zmura, S. Deng, T. Lappas, S. Thorpe, and R. Srinivasan. "Toward EEG sensing of imagined speech," *Human Computer Interactions and New Trends*, pp. 40-48, 2009.

[2] EGI. (2008). [Online]. Available: http://www.egi.com/index.php

[3] BCI Competition III. (2005). [Online]. Available: http://www.bbci.de/competition/iii/

[4] D. Donoho, "Compressed sensing," *IEEE Trans. Info. Theory*, vol. 52, 1289–1306, 2006.

[5] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation". *IEEE Trans. on Signal Processing*, 54(11):4311, 2006.

[6] H. Lee, A. Battle, R. Raina, and A.Y. Ng, "Efficient sparse coding algorithms". *Advances in neural information processing systems*, 19:801, 2007.

[7] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online learning for matrix factorization and sparse coding". *Journal of Machine Learning Research*, 11(1):19–60, 2010.

[8] R. Baraniuk, "Compressive sensing". *Lecture notes in IEEE Signal Processing magazine*, 24(4):118–120, 2007.

[9] K. Koh, S. Kim, and S. Boyd. l1 ls: A Matlab Solver for Large-Scale 1- Regularized Least Squares Problems. 2008.

[10] E.R. Kandel, J.H. Schwartz, T.M. Jessell, S. Mack, and J. Dodd. *Principles of neural science*. Elsevier New York, 1985.

[11] J. S. Lim and A. V. Oppenheim. *Advanced Topics in Signal Processing.* Englewood Cliffs, NJ: Prentice Hall, pp. 87-89, 1988.

[12] S. M. Kay. *Modern Spectral Estimation: Theory and Application.* Englewood Cliffs, NJ: Prentice Hall, pp. 228-230, 1988.

[13] Chih-Chung Chang and Chih-Jen Lin, LIBSVM : a library for support vector machines, 2001. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm