University of Alberta

An Investigation of the Use of EEG Phase
in Groupwise Classification

by

A. Matthew Landals ©

A thesis submitted to the Faculty of Graduate Studies and Research in partial
fulfillment of the requirements for the degree of Master of Science

in

Statistics

Department of Mathematical Sciences

Edmonton, Alberta

Fall 2000

0-612-59827-6

University of Alberta

Library Release Form

Name of Author: A. Matthew Landals

Title of Thesis: An Investigation of the Use of EEG Phase in Groupwise Classification

Degree: Master of Science

Year this Degree Granted: 2000

Matthew Landals
A. Matthew Landals
13747 Woodcroft Avenue
Edmonton, CANADA
T5M 3M3

September 25 2000
Date

UNIVERSITY OF ALBERTA

Faculty of Graduate Studies and Research

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled **An investigation of the use of EEG phase in groupwise classification** submitted by **A. Matthew Landals** in partial fulfillment of the requirements for the degree of **Master of Science** in Statistics.

Dr. B.A. Schmuland (Supervisor)

Zoltan J. Koles

Dr. Zoltan J. Koles (Co-supervisor)
(Biomedical Engineering)

Dr. N.G.N. Prasad (Chair)

John C. Lind

Dr. John C. Lind (Alberta Hospital)

Date: September 15 2000

Abstract

Using phase of EEG data to classify control subjects groupwise according to four tasks was investigated. Phase alone proved unsuccessful, but phase/frequency correlation, slope and direction of channel pairs receiving similar but lagged signals were more promising. Individual EEG data arrays (20 one-second epochs by 43 channels by 256 measurements per second) were Fourier transformed. Channel pair covariances were computed, smoothed and stored as phase arrays. Data was reduced to a manageable form useful for discrimination by finding highest correlation channel pairs for subject phase arrays and for task phase arrays. Of various categorical and numerical methods investigated and tested, only one using the correlation/slope vectors of highest correlation task channel pairs had marginal success in that it produced four-way discrimination where P(Type I error)<(1-P(Type II error)). Low between-task phase/frequency variation inhibited complete, reliable, efficient, four-way task identification.

Acknowledgements

I am very grateful for the help of my supervisors, Dr. Byron Schmuland and Dr. Zoltan Koles, in the research and writing of this thesis.

Dr. Schmuland provided financial support. He was enthusiastic in testing many possible research directions and very helpful in directing me to numerical methodologies. Dr. Schmuland's help in setting up mathematical proofs was greatly appreciated. Moreover, he gave invaluable assistance in producing this document and in critically reading it.

Dr. Koles provided crucial guidance by directing me to many pivotal aspects of the research. I would like to thank him for giving me the opportunity to work on this project. I thoroughly enjoyed it.

I am also indebted to Dr. John Lind. Dr. Lind allowed me to use data from his research and generously gave of his time to show me the data collecting procedure. He contributed many ideas through discussion and was always ready to provide any information I needed.

I would also like to thank Dr. Prasad for his critical appraisal of the thesis and his helpful suggestions for improving it.

Lastly, I thank Ion Buicliu for his always willing help in the Biomedical Engineering computer laboratory.

Contents

List of Tables

List of Figures

List of Symbols and Abbreviations

V – raw EEG matrix

t – time within one second

FT – Fourier transform operation

F – Fourier transform matrix, i.e. VF is V multiplied by the Fourier transform matrix

ω – frequency (Hz)

N – length of series

$\hat{V}=FT(V)$ – V in the frequency domain

W – lag of V

$\hat{W}=FT(W)$ – W in the frequency domain

s – one-second epoch

a, b, c, $\ldots$ – arbitrary channels

C – covariance array using only epoch aggregation smoothing

CMA – covariance array with moving average smoothing but no principal component analysis smoothing

PMA – phase(CMA) – phase array with moving average smoothing but no principal component analysis

PCA – principal component analysis

T – 43×43 matrix of complex covariances when doing principal component analysis

CPD – covariance array after all smoothing, including principal component analysis

PPD – phase(CPD) – phase array with principal component analysis smoothing

MSR – mean regression squares

MSE – mean error squares in linear regression

SSR – total regression squares

SSE – total error squares in linear regression

SST – total sums of squares, explained or unexplained

CBP – task covariance array

PBP – task phase array – $PBP = \text{phase}(CBP)$

A, B – arbitrary tasks

$\overline{x}$ – mean vector

Σ – covariance matrix

DL – dot localization task

WF – word finding task

EO – eyes open task

EC – eyes closed task

Rel. – reliability

Eff. – efficiency

Chapter 1

Introduction

Psychiatrists would find it valuable to have more refined and rigorous quantitative methods for analyzing their patients' electroencephalograms or EEGs to determine the specific nature of psychiatric impairment. Discrimination using EEG data could yield a more efficient and accurate method of diagnosis than the longer process of observing patient behavior and might be employed in diagnosis before symptoms become obvious. This, in turn, would be beneficial in administering medication and monitoring its results. EEGs have an advantage over other techniques of assessing brain function, such as Functional Magnetic Resonance Imaging (fMRI), since they can measure rapid changes in brain activity.

1.1 Use of EEG Data in Discriminant Analysis

One approach to discriminant analysis has been to compare the EEGs of known non-psychiatrically impaired controls and patients with known psychiatric impairment when they perform standard tasks. Various magnitude methodologies like power spectra, cross-spectra and power ratios, which use EEG data in the frequency domain, have then been used to classify the EEGs of controls and patients. The classification rules which are developed from these known cases are then applied to unknown cases.

Good examples come from studies of the EEGs of known ischemic and non-ischemic subjects performing standard tasks, where classification involved analyzing the relative spectral powers of the left and right sides of the brain. Pfurtscheller, Auer, and Hoprumer (1984) [11] calculated an asymmetry index, a weighted sum of three asymmetry ratios, for every subject. A distance function was chosen so that a positive value could be ascribed to normal subjects and a negative value to ischemic patients. Van Huffelen, Poortvliet, and van der Wulp (1984) [17] performed discriminant analysis using untransformed asymmetry ratios. It was concluded that ischemia was present on that side of the brain which had significantly lower spectral power than the other and that ischemia was absent if the two sides had almost equal power. All subjects were assigned a classification as normal or ischemic. However, while the sensitivity or probability of correctly classifying a control as non-ischemic was 95%, the specificity or probability of correctly classifying an ischemic patient was only 55%.

Lind, Koles, Flor-Henry, and Soong (1997) used the EEGs of 33 right-handed female control subjects to discriminate among three tasks, dot localization, word finding and eyes closed. Once data was Fourier transformed to the frequency domain, discrimination was performed using the quadratic discrimination function (Morrison, 1990) [10]. In three discrimination methods, two using cross-spectra and one using power spectra, the basis for classification was the distance of each subject's EEG from the centre of the task group. Three discrimination scores were computed to develop classification rules. Although discrimination scores were used to separate the three tasks simultaneously, eyes closed was readily identified by a large alpha band component. All EEGs were classified independently of a reliability standard. The use of cross-spectra was very successful, especially where 41 EEG channels were used, since the proportions of correct classifications were mostly above 80%. Using 16 EEG channels also worked well, except that less than 80% of word finding and eyes closed subjects were correctly classified. Where power spectra were used, all word finding subjects' EEGs were misclassified as either dot localization or eyes closed. The last two results show how making a discrimination decision on every subject can lead to high probabilities of misclassification.

Using a data set consisting of the EEGs of 69 right-handed female controls, Lind, Flor-Henry, and Koles (1999) [8] employed spectral and cross-spectral analysis to discriminate between two active tasks (dot localization and word finding) and between two passive tasks (eyes open and eyes closed). A classical likelihood ratio approach was used as the general framework for discrimination methodology (Rao, 1973) [13]. First, data was smoothed using a 50% Hamming taper and correlated effects were removed using a digital filter (Brillinger, 1981) [1]. Then, as an alternative to analyzing 43 power spectra (one for each EEG channel) and 903 cross-spectra (one for each channel pair (a, b) where $a < b$) (Priestly, 1981) [12], a quadratic discrimination function was used. The cross-spectral matrices were factored into complex spatial patterns. The spatial patterns that accounted for maximum EEG variance in one task and minimum variance in the other were used to calculate scores for each subject (Morrison, 1990) [10]. Discrimination was based on these scores by choosing a decision boundary to maximize the proportion of correct classifications. A classification decision was made for all subjects independently of a reliability standard, and while the proportion of correct classifications was 80% on average, it fell into the 70's in some cases. This study did not distinguish between active versus passive tasks and so did not discriminate among all four tasks at once.

1.2 Extending EEG-based Discriminant Analysis

When only the magnitudes of EEG signals in the frequency domain are used, only half the information EEG data has to offer is used. The other half, phase, has not been applied to discriminant analysis using EEGs.

In this project, the goal was to investigate the use of phase in discrimination by attempting to perform groupwise classification while setting a confidence standard so as to avoid high misclassification probabilities.

The thesis proposed is: it is possible to produce a phase-based methodology from EEG data which gives rules to correctly classify subjects as performing one of four standard tasks while maintaining a reliability of at least 0.8 and an efficiency of at least 0.5.

This research is part of a larger project employing both EEG magnitude and phase to discriminate between types of psychiatrically impaired patients which is being conducted at the Clinical Diagnostics and Research Centre, Alberta Hospital, Edmonton, and the Department of Biomedical Engineering, University of Alberta. The conventions of the larger study are followed.

Chapter 2

EEGs: Data, Biophysics and Model

EEGs record the potential differences of brain currents received by electrodes attached to the scalp. While the actual electrical patterns produced in the brain are unknown, physics gives us enough information to establish a relationship between unknown brain current and potential difference received by electrodes and to develop a matrix model relating the two. The model gives confirmation that voltage produced in the brain is linearly correlated with voltage received by electrodes and that random variation or noise in voltage is normally distributed. The electrical pattern in the brain is highly determined by combinations of gender, handedness, psychiatric state and task-oriented brain function. If three out of these four factors are kept constant, then, in theory, differences in EEGs would be highly explained by differences in the fourth factor. There is, however, considerable noise due to unknown determinants of electrical activity.

2.1 Recording EEGs

Electrodes are placed on the scalp according to the standard electrode positions described in the American Electroencephalographic Society guidelines (Sharbrough et al, 1990) [15]. A total of 48 electrode channels are attached to the scalp, but only 43 of these are EEG channels. The others are reserved for special functions such as electromyography (EMG) which records muscle artifact and electrocardiography (EKG) which records heart activity. The EEG channels, numbered 1 to 43, reference specific EEG sites and these are listed in Appendix A.

EEG data is measured discretely, being segregated into one-second time periods or epochs. In every one-second epoch, the potential difference at the 48 electrodes relative to a reference electrode is recorded to the nearest microvolt 256 times.

2.2 Project Data

The data for this project was gathered at the Alberta Hospital, Edmonton (Lind, pers. comm.) [9]. It consists of the EEGs of 88 female, right-handed, non-psychiatrically impaired controls, each performing one of four standard tasks. Twenty-two control subjects were assigned dot localization, 20 were given word finding, 23 performed the eyes open task and 23 performed the eyes closed task.

Dot localization and word finding, which make the subject think, are active tasks. For dot localization, the subject is presented with a card on which there are two rectangles of equal size and shape. In one rectangle there are two dots, while in the other there are several numbers. The subject is asked which two numbers the dots would cover if one rectangle was superimposed on the other. For word finding, the subject is given a dictionary definition and asked to think of a word to fit that definition.

Eyes open and eyes closed do not require thought and are passive tasks. For eyes open, the subject is asked to sit passively with her eyes open. For eyes closed, the subject is asked to sit passively with her eyes closed but to remain awake.

2.3 Data Collection Conventions

Data collection follows standards set by the Alberta Hospital, Edmonton (Lind, pers. comm.) [9].

(i) Active tasks have a cycle in which the subject is given the question, thinks about the response and then presses a button just before responding. Only the time period between the subject getting the question and pressing the button, that is, the time the subject is thinking, is counted. One second of this thinking period is selected for analysis. If the subject gives the wrong answer (not even half right), the thinking period must be discarded because there is no proof of relevant thought.

(ii) There is a limit on how much muscle artifact can be present in the data to be analyzed. Muscle artifact indicates a lack of subject relaxation and consists of high amplitude, high frequency EMG. Only periods where the muscle artifact is low can be selected.

(iii) For a subject's EEG to be analyzed, there must be a certain number of one-second time periods or epochs which satisfy rules (i) and (ii). There must be at least 20 such epochs for control subjects and 10 such epochs for psychiatrically impaired patients.

(iv) Because it can be difficult for a subject's EEG to satisfy rules (i), (ii) and (iii), the one-second time periods need not be adjacent. Because any amount of time can elapse between chosen periods, it is assumed that the one-second epochs are independent.

(v) To satisfy rules (iii) and (iv) for active tasks, the subject is given a series of more than 20 different tests of the assigned task. This allows 20 different one-second periods or epochs of EEG data to be extracted from the total number of tests. Subjects performing a passive task remain in the passive state continuously for several minutes. Twenty different one-second epochs of EEG data are selected from this total period.

(vi) Potential differences are measured in microvolts rounded to the nearest integer, resulting in up to four significant digits of accuracy. After this, raw data can be stored as 12-bit signed integers (-2048 to 2047).

(vii) It is usual for the Alberta Hospital to remove means and trends in potential differences from every epoch because they contain no useful information for data analysis. While this was not done for this project, it will have no effect on the results. Demeaning and detrending are high pass filters affecting only the frequencies of 0 Hz and 1 Hz. Changes in these low frequencies of EEGs are almost always due to changes in surrounding phenomena rather than changes in brain activity.

2.4 Biophysical Properties of EEGs

Brain current comes from cells in the gray matter which are responsible for performing brain functions. It is in the gray matter that cell groupings called dipoles produce polarized current where positive and negative charges flow in a definite direction. In inactive regions of gray matter, net current is essentially zero. This is because current sources are randomly aligned due to the non-stationarity of the dipoles. In active regions, a non-zero current is produced because dipoles are highly stationary and current sources line up.

Electrical current flows out of the gray matter and through highly conductive white matter to other gray matter tissues or to the body to make it function. Voltage or electrical potential is significantly reduced by the time it reaches the scalp due to the high resistivity of the skull. However, this voltage is still non-zero and electrodes attached to the scalp can receive signals.

Electrodes cannot measure scalp voltage directly so differences in voltages are measured. EEGs record differences between the voltage received by scalp electrodes and the voltage received by a reference electrode attached in an area where voltage is likely to remain constant. These voltage differences are called potential differences.

2.5 The EEG Model

The bioelectrical physics behind the EEG leads to the following model: $V = MS$ where V, M, S are matrices (Koles, pers. comm.) [6].

S is the source matrix whose (k,t)th entry $S_{k,t}$ represents the voltage produced by the kth dipole during the tth 1/256 of a one-second epoch. Each such element consists of deterministic signal plus random noise, that is, $S_{k,t} = \sigma_{k,t} + \epsilon_{k,t}$. The number of rows in S is the total number of dipoles in the brain, a figure that is unknown but effectively infinite. The number of columns in S is the number of recordings per epoch which is 256.

M is a deterministic matrix whose (i,k)th entry $M_{i,k}$ measures the contribution from the kth dipole to the voltage recorded at the ith electrode channel. The number of rows in M is the number of electrode channels, 48, and the number of columns is the number of dipoles in the brain. The matrix M reflects all resistive and smearing effects of the skull. If smearing was great enough, all values of M would be approximately equal. However, EEGs do detect voltage differences between different regions of the brain and so it can be deduced that smearing effects do not prevent $M_{i,k}$ from being significantly larger when dipole k is close to electrode i. Therefore, there is a reasonably strong linear correlation between potential differences received by electrodes and the voltage of electricity produced by nearby regions inside the brain.

Multiplying M and S gives V which is the matrix containing raw EEG data for a one-second epoch. That is, V is a 48×256 matrix whose (i,t)th entry $V_{i,t}$ is the voltage (in microvolts) recorded at the ith electrode channel during the tth 1/256 of the epoch. We have

$$V_{it} = \Sigma_k M_{ik} S_{kt} = \Sigma_k M_{ik}(\sigma_{kt} + \epsilon_{kt}) = \underbrace{\Sigma_k M_{ik}\sigma_{kt}}_{\text{signal}} + \underbrace{\Sigma_k M_{ik}\epsilon_{kt}}_{\text{noise}}. \qquad (2.1)$$

The number of dipoles in the gray matter is very large in relation to the number of electrodes and so the total noise consists of the sum of many independent random values. Therefore, the central limit theorem allows us to assume that the noise in the V matrix is normally distributed no matter how far from normal the distribution of the noise in S is.

To emphasize the special role played by the time variable, we sometimes write $V(t)$ for the column vector $(V_{1t}, \ldots, V_{48t})'$, and $V_i(t)$ for V_{it}.

A subject's EEG record consists of 20 one-second epochs, 48 channels and 256 signal measurements per second. The 20 one-second epochs are not necessarily adjacent and so are assumed to be independent. The raw data is stored as a $20 \times 48 \times 256$ array, which, for analysis purposes, can be considered as 20 independent V matrices.

The EEG data to be analyzed becomes a $20 \times 43 \times 256$ array when the non-EEG channels are removed.

Here is an example of all 256 raw EEG readings (in microvolts) from the first one-second epoch of one channel (29) of a dot localization subject. The readings are in sequence from left to right and are plotted in Figure 2.1. Wherever possible, this same data set, the first on a randomly ordered list of all subjects, is used throughout for illustrative purposes and is referred to as Dot Localization 1.

158	145	126	115	102	87	60	24	−4	−23	−34	−22
11	47	61	43	29	30	34	62	100	113	126	151
129	59	−18	−93	−162	−193	−173	−123	−58	3	44	57
49	30	21	9	−11	−13	−3	−5	6	29	35	38
55	65	67	60	14	−40	−60	−72	−56	−2	18	−4
−5	−13	−53	−53	−4	7	−14	−13	−29	−73	−78	−39
−19	−17	−5	4	10	19	7	−21	−48	−67	−62	−38
−27	−21	4	17	9	21	45	46	26	8	−39	−90
−100	−76	−47	−20	−12	−32	−59	−68	−43	2	34	40
35	23	−4	−33	−49	−57	−78	−97	−91	−67	−31	31
93	122	122	104	62	25	12	−7	−43	−83	−128	−151
−128	−87	−73	−78	−72	−64	−61	−51	−40	−37	−21	21
63	87	88	73	46	13	−11	4	44	66	79	86
69	41	29	26	24	9	−31	−64	−74	−70	−38	15
37	44	69	73	34	−4	−35	−75	−103	−104	−109	−106
−69	−26	−12	4	24	26	41	79	99	117	155	167
156	162	151	96	24	−39	−95	−119	−109	−100	−92	−64
−34	−26	−34	−38	−31	4	52	91	115	119	92	49
21	15	31	53	52	32	17	11	5	−13	−37	−49
−45	−34	−20	−9	−20	−43	−46	−25	−1	30	55	51
29	12	−11	−35	−46	−55	−63	−66	−75	−58	6	77
105	88	56	34								

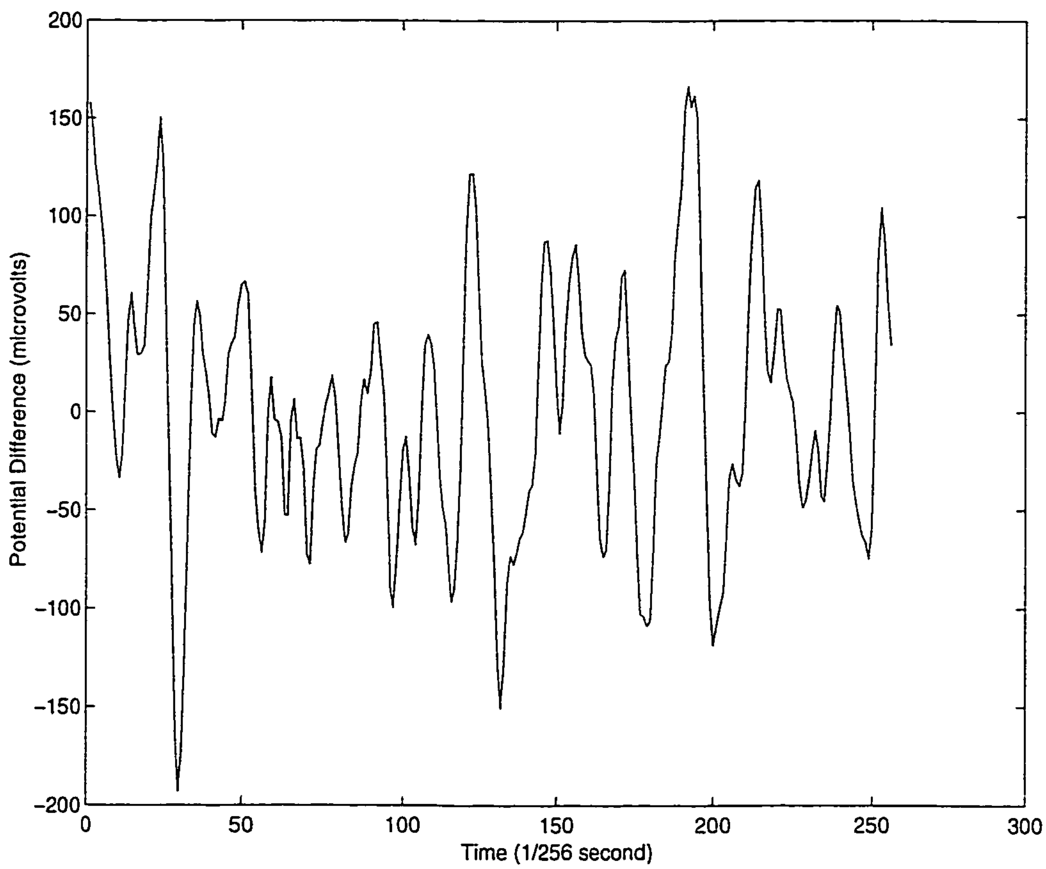

Figure 2.1: Time Series Plot for a One-Second Epoch
The data is the first epoch from channel 29 of Dot Localization 1.

Chapter 3

Fourier Transform, EEG Signal and Normality of Noise

Employing phase as a tool for discriminant analysis first involves Fourier transforming the EEG data from the time domain where numbers are real to the frequency domain where numbers are complex and thus have phase. Noise, along with deterministic signal, is present in all EEG data. Before proceeding with discriminant analysis, normality must be established so that tests of hypotheses can be carried out to prove that data contains signal and is not entirely noise.

3.1 Fourier Transform

A subject's EEG data array is three-dimensional, with a between one-second time dimension, a within one-second time dimension and a channel dimension. Since the 20 one-second epochs are independent, the Fourier transform must be applied to the within one-second time dimension and not to the between one-second time dimension.

Each epoch is considered to be an independent replication of an observation of a 43-dimensional vector-valued stationary time series, $V_t = [V_1(t), V_2(t), \ldots, V_{43}(t)]$ at the time points ($t = 0, \ldots, 255$). Calculating the discrete Fourier transform of the vector V provides a 43×1 complex-valued vector. The Fourier transform formula is $f_\omega = (1/256)\sum_{t=0}^{255} V_t \exp(-i2\pi\omega t/256)$ at frequencies $\omega = 0, \ldots, 255$. The Fourier transform can also be expressed in matrix notation as $FT(V) = VF$ where F is a 256×256 matrix and $F_{\omega t} = (1/256)\exp(-i2\pi\omega t/256)$.

If the Fourier transform was performed directly and the length of the vector was N, N frequency domain numbers would have to be calculated and computing each one would involve performing N exponents. Therefore, the direct Fourier transform is a $\theta(N^2)$ algorithm.

The more efficient fast Fourier transform algorithm splits the array into the odd and even halves, Fourier transforms each half recursively and combines the halves by taking the average. The result is a $\theta(N \log N)$ algorithm. A description of the properties of the Fourier transform and how the fast Fourier transform algorithm works is found in Gonzalez and Wintz (1977) [4].

The production of 256 unrelated complex numbers from 256 real numbers via the Fourier transform would double the amount of information since complex numbers have both a real component and an imaginary one. Since this is mathematically impossible, half the resulting complex numbers must be related to the other half. As it happens, the upper 128 frequencies are complex conjugates of the lower 128 frequencies. This can be proved directly from the Fourier transform formula.

Let $\hat{V}$ be the Fourier transform of a time domain series, V, and ω be a frequency.

Then,

$$
\begin{aligned}
\hat{V}(256-\omega) &= \frac{1}{256}\sum_{t=0}^{255} V(t)\exp\left(\frac{-i2\pi(256-\omega)t}{256}\right) \\
&= \frac{1}{256}\sum_{t=0}^{255} V(t)\exp\left(\frac{i2\pi\omega t}{256} - i2\pi t\right) \\
&= \frac{1}{256}\sum_{t=0}^{255} \overline{V(t)\exp\left(\frac{-i2\pi\omega t}{256}\right)} \\
&= \overline{\hat{V}(\omega)}.
\end{aligned}
$$

To illustrate, the real and imaginary parts of the Fourier transformed data taken from the first one-second epoch of channel 29 of Dot Localization 1 are shown in Figures 3.1 and 3.2 respectively.

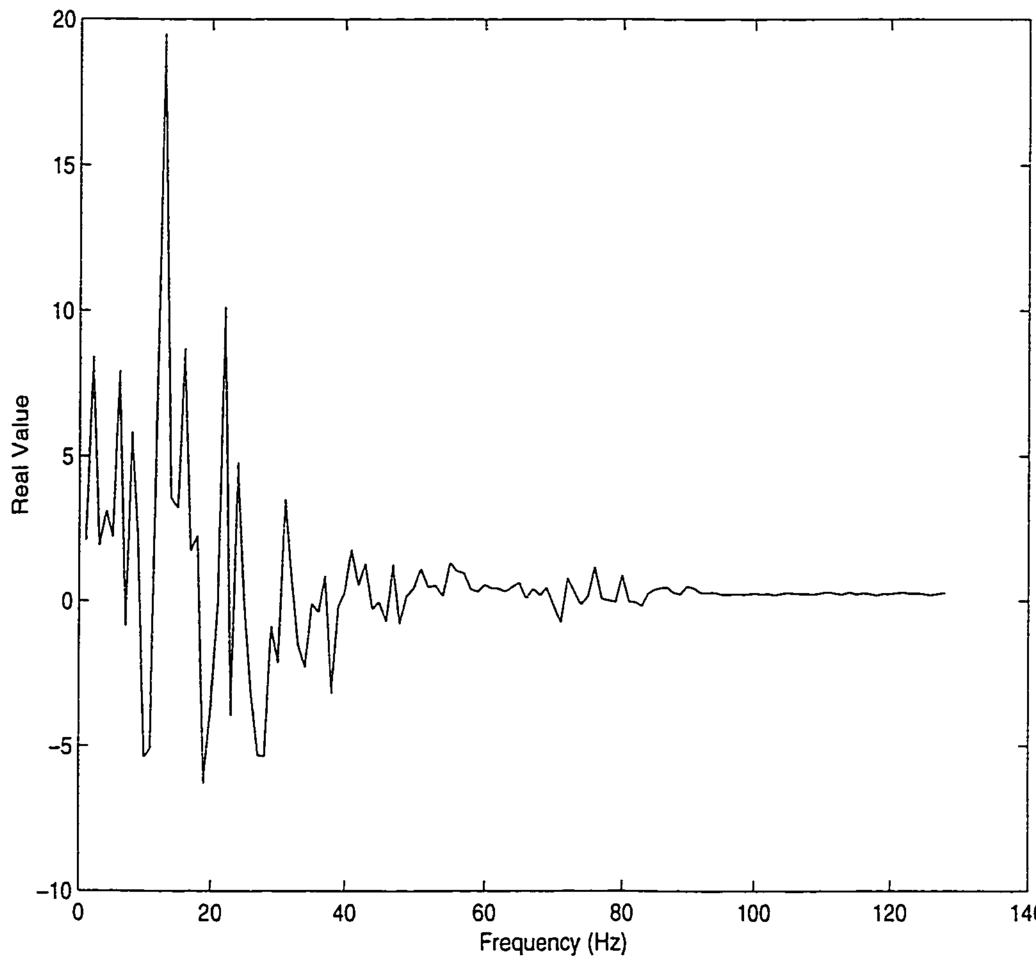

Figure 3.1: Real Part of a Fourier Transformed One-Second Epoch
This is the first epoch for channel 29 of Dot Localization 1. Only frequencies 0 to 127 Hz are plotted since the upper 128 frequencies are complex conjugates of the lower.

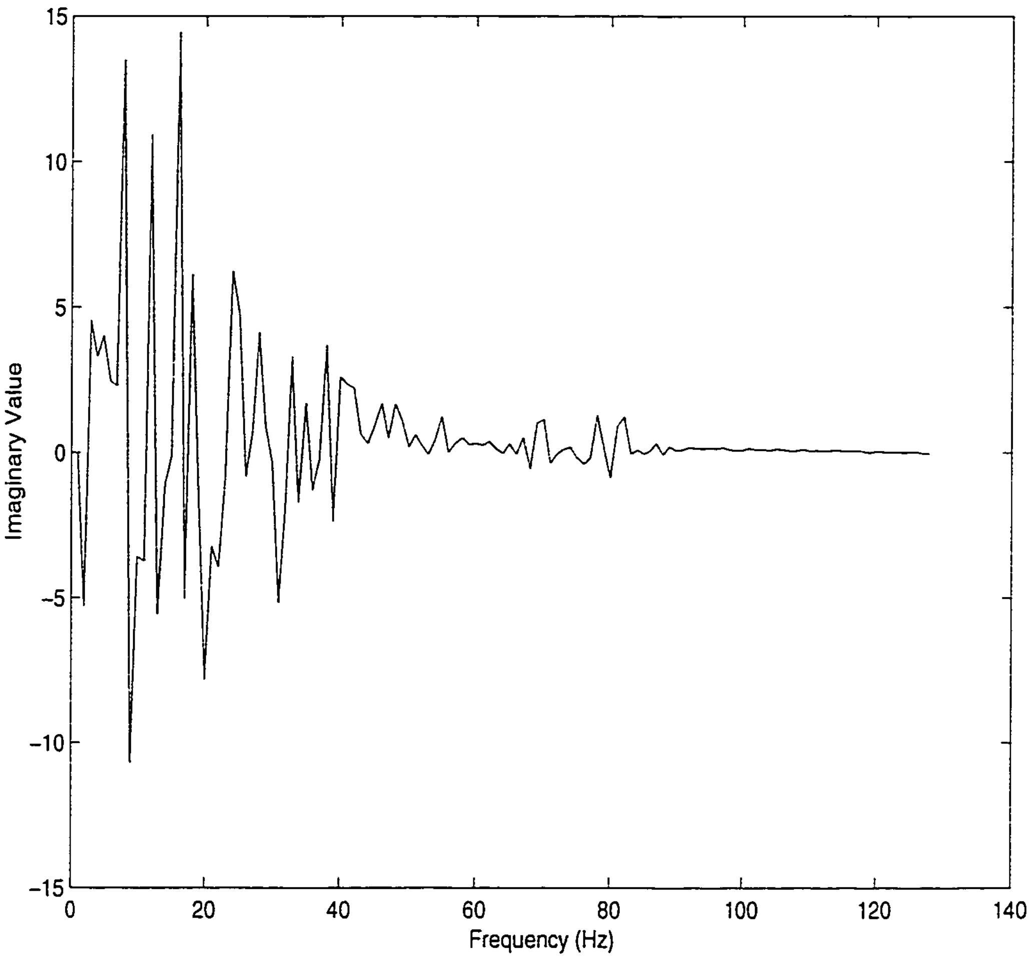

Figure 3.2: Imaginary Part of a Fourier Transformed One-Second Epoch
Again, this is the first epoch for channel 29 of Dot Localization 1 and only 0 to 127 Hz are plotted as the upper 128 frequencies are complex conjugates of the lower.

3.2 Phase

After Fourier transformation, the data for one subject is a $20 \times 43 \times 256$ array of complex numbers from which phase can be calculated. The phase of a complex number is the inverse tangent of the imaginary part divided by the real part, that is, every complex number $z = \Re(z) + i\Im(z)$ can be expressed as $z = |z|e^{i\theta}$, where $|z| = \sqrt{\Re(z)^2 + \Im(z)^2}$ is the modulus of z, and θ =arctan($\Im(z)/\Re(z)$) is the phase of z (Brown and Churchill, 1996) [2].

In order to reduce phase noise and interpret phase properly, it is important to note that phases are only congruent modulo 2π. This means that if 2π radians are added or subtracted from a phase, it would be the same. For example, 0 radians = 2π radians = 4π radians and so on. Therefore, when handling phase, care must be taken to unwrap the data (add or subtract 2π radians where appropriate) before drawing any conclusions.

3.3 Normality

Normality of noise in EEG data must be established before and after Fourier transform to help construct tests of hypotheses which prove that EEG data contains deterministic signal and not just noise.

The noise of every element in V, a one-second epoch of EEG data before Fourier transform, is the sum of many independent random variables (Section 2.5). Therefore, the central limit theorem gives a theoretical reason for assuming normality of noise.

To illustrate the normality of V in practice, a normal probability plot of potential differences can be plotted (Figure 3.3). Data plotted is a subset of Dot Localization 1's EEG consisting of one EEG channel (29) and all 20 one-second epochs, but only 1/256 of each epoch. The last is so that change in the deterministic EEG signal, which depends on time and channel, cannot mask the normality of the noise. Allowing for the illusion of a sinusoidal pattern due to the small sample size, the closeness of the points to the line equating observed and expected cumulative probability is evidence of the normality of the noise.

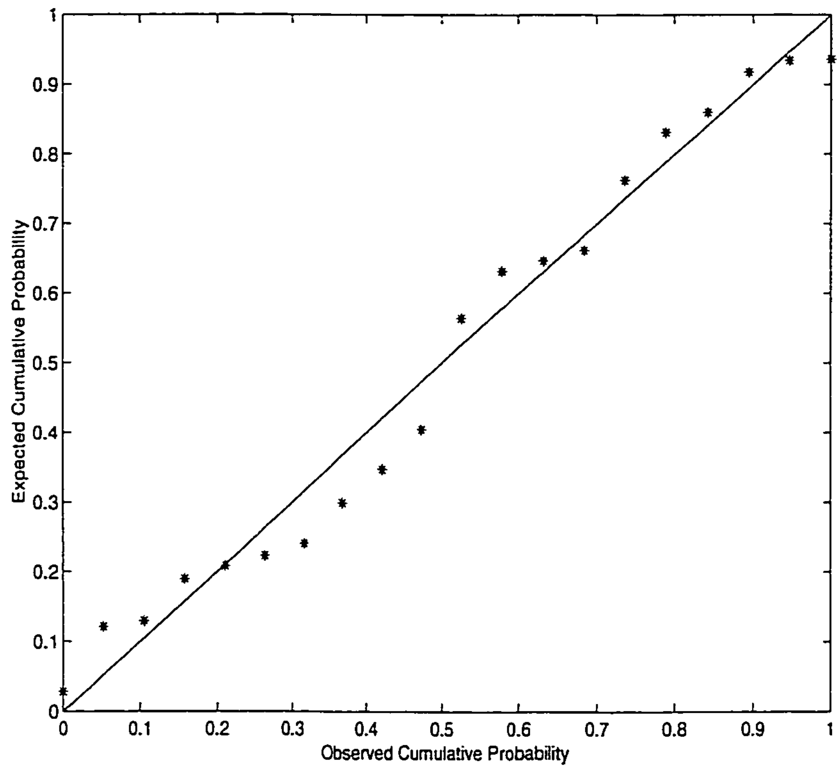

Figure 3.3: Normality Plot of Potential Differences from One Subject's EEG Data plotted is for Dot Localization 1 and consists of one channel (29) and all 20 epochs, but only 1/256 of each epoch. Such a data subset eliminates deterministic signal which depends on time and channel, but leaves the random variation. The small sample size produces the illusion of a sinusoidal pattern.

Since the Fourier transform is a linear transformation, normality carries over to the frequency domain. Each column of the Fourier transformed matrix VF has the complex normal distribution with mean $\mu(\omega) = (1/256)\sum_{t=0}^{255} E(V_t)\exp(-i2\pi\omega t/256)$. As is usual in regression analysis, we assume that the columns of V are independent with a common covariance matrix Σ. In this case, the covariance of each column of VF is $E(VF(\omega)[VF(\omega)]') = (1/256)\Sigma$.

3.4 Normality Under H_0: Noise and No Signal

The normality of EEG noise can be used to test the null hypothesis that the data is only noise and contains no signal.

Suppose the data contains only noise and no signal. Then, in a $20 \times 43 \times 256$ array, all matrices V will be identically as well as normally and independently distributed, even after Fourier transform. Therefore, the matrix elements can be converted to the standard normal distribution by normalizing across the between one-second time dimension. Considering elementary statistical theory as in Freund (1992) [3], the following will be true when H_0 is true.

(i) The magnitude of each element, except for elements representing frequency 0, will have χ^2_2.

(ii) The magnitude of each element representing 0 frequency will have χ^2_1.

(iii) Except for frequency 0, the square of the imaginary part divided by the square of the real part will have $F_{1,1}$. The real and imaginary parts will be uncorrelated and hence independent since they are normally distributed.

(iv) The complex number plane can be broken down into the four quadrants. If neither the real part nor the imaginary part of any complex number is 0, then each point will fall into each of the four quadrants with equal probability. Furthermore, the number of points in any quadrant will have binomial$(0.25, n)$ distribution where n is the number of points.

(v) The complex number plane can be broken down into octants by dividing each of the four quadrants into two equal halves. The number of points in each octant will have binomial$(0.125, n)$ distribution.

To test statements (i), (ii) and (iii), empirical evaluations based on χ^2 and F distributions were performed. One Fourier transformed, standardized epoch of Dot Localization 1's EEG was tested. Results concluded that H_0 is false since the P-value was less than 0.01.

Statements (iv) and (v) were concluded false because the data points do not fall into the four quadrants or eight octants with equal probability. Figure 3.4 shows an example of how the points in a standardized V matrix fall mainly into the first and fourth quadrants.

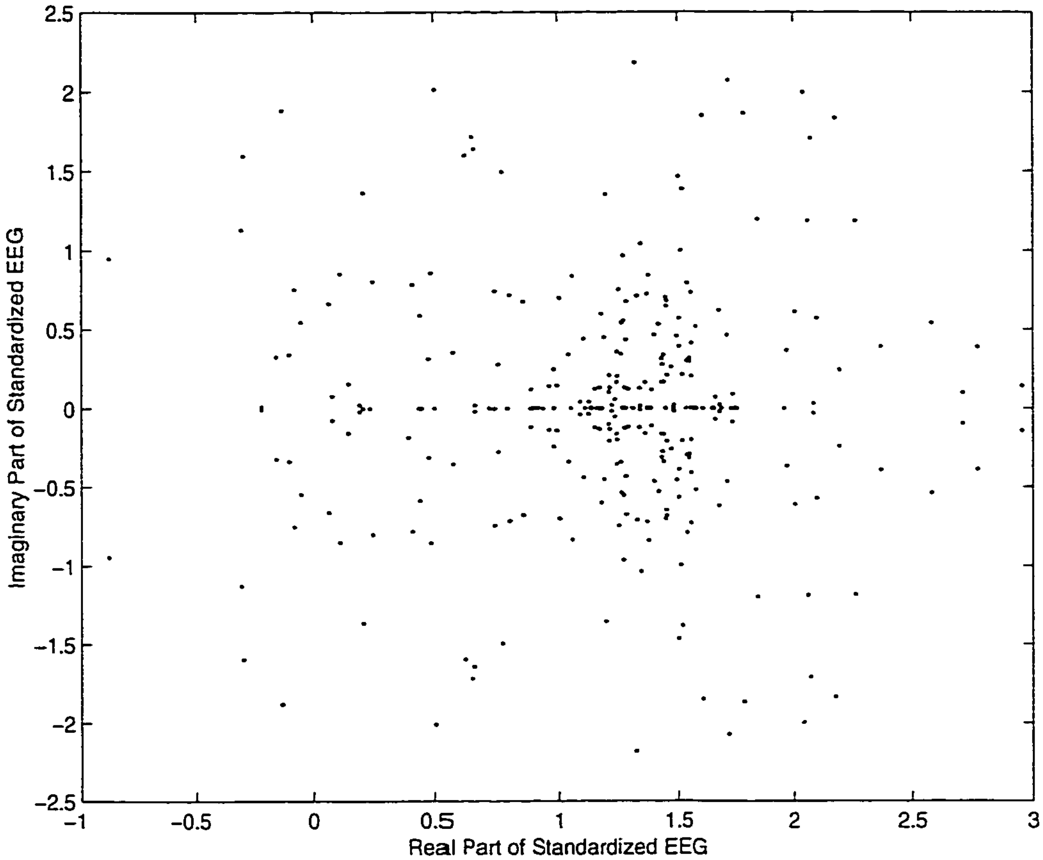

Figure 3.4: Scatterplot of Real and Imaginary Components of EEG Data after Fourier Transform and Standardization

Data plotted consists of the values of a 43×256 matrix taken from Dot Localization 1's EEG after Fourier transformation and standardization. The first and fourth quadrants contain more data points than the second and third. Since the points do not fall into each of the four quadrants or eight octants with equal probability, it is evidence against the null hypothesis that the EEG data is only noise.

Since all statements are false, it is evidence that the null hypothesis is false. Therefore, EEG data contains signal and not just noise. This justifies its use in discriminant analysis.

Chapter 4

EEG Phase and Channel Pair Phase/Frequency

The Fourier transform converts each subject's EEG data to a $20 \times 43 \times 256$ array of complex numbers which contain phase. When first considering an alternative method to magnitude-based discriminant analysis, it was thought that using only the phase of EEG data, after Fourier transform alone, might produce criteria to classify subjects according to task. Various methods of using phase alone, directly or indirectly, were consequently investigated and discarded. Eventually, the lead-lag theory of pairs of EEG signals led to considering differences in the phase/frequency relationships of channel pairs as a method of discrimination.

4.1 Using Phase Alone

Taking the Fourier transformed arrays without any further data processing and pursuing methods of using the phase of complex numbers alone proved fruitless. Three of these, phase clustering, considering sums of real squares and imaginary squares, and normalization of epochs, are worth mentioning.

4.1.1 Phase Clustering

When testing EEG data for presence of signal, it was found that Fourier transformed, standardized EEG data points did not fall into the four quadrants of the complex number plane with equal probability. Therefore, it was reasoned that data points might fall into different quadrants for different tasks. The number of data points falling into each quadrant was determined for individual epochs of each subject's EEG and those quadrants which had significantly more data points than others were noted. However, those quadrants with the most data points varied more between epochs of the same subject's EEG than they varied between tasks.

4.1.2 Sums of Real Squares and Imaginary Squares

Another idea was that the real part of the Fourier transformed EEG data might contain more information than the imaginary part or vice versa. Therefore, phase was used indirectly by trying to find differences in the sums of squares of either the real parts or the imaginary parts. A search was performed to find channels where the sums of either real squares or imaginary squares, computed across the within one-second time dimension, had high variation between tasks but low variation between epochs of the same subject's EEG. This was abandoned when it was found that no matter what the channel, there was too much variation between subjects performing the same task.

4.1.3 Normalization

Because total EEG magnitude depends mainly on skull thicknesses, an attempt was made to improve the indirect use of phase by normalizing EEG data so that every subject would have the same alpha (frequencies 8 to 13 Hz) power. Normalization was applied to every epoch of every subject's Fourier transformed EEG data. However, when the sums of real squares and imaginary squares were reexamined after normalization, there was no improvement in discrimination.

4.2 Using Channel Pair Phase/Frequency Relationships

Phase alone probably fails to discriminate because it is hampered by the random absolute timing of EEG signals which causes too much within-task and within-subject variation. One method of considering relative timing is to look at the relationship between phase and frequency for pairs of channels.

A rationale for examining channel pair phase/frequency relationships as a method of discrimination comes from the lead-lag theory of pairs of EEG signals. It is hypothesized that when the brain performs a function, certain activated regions lag behind other activated regions. This hypothesis means that there exist pairs of electrodes receiving similar signals, with the signal received by one channel lagging behind that of the other. The presence of such a lag promises that once the data is processed beyond the Fourier transform, it is possible to find pairs of channels that yield linear relationships between phase and frequency, preferably ones that differ between tasks. A discussion of signal delay versus phase is found in Hannan (1983) [5].

4.3 Channel Pair Phase/Frequency from Time Lags

We now process every series beyond the Fourier transform to get a phase/frequency relationship. Let $\hat{V}(\omega)$ and $\hat{W}(\omega)$ be the Fourier transform of V and W respectively. Then covariances are computed by multiplying $\hat{V}(\omega)\overline{\hat{W}(\omega)}$ for every ω. Taking the phase of the resulting complex numbers gives a phase/frequency relationship, $\text{Phase}(\hat{V}(\omega)\overline{\hat{W}(\omega)})$ versus ω.

Suppose two channels receive EEG signals whose only difference is a time lag. Then the phase/frequency relationship computed for that pair will be a perfectly straight line. We can illustrate the principle by proving the special case of a lag of 1/256 of an epoch.

In the time domain, let $V = V_0, V_1, \ldots, V_{254}, V_{255}$ and define W to be the V signal at lag 1, that is, $W = V_{255}, V_0, \ldots, V_{253}, V_{254}$. Setting $N_{\omega t} = \exp(-i2\pi t\omega/N)$, we can write the Fourier transform of W as $N_{\omega(t-1)}\exp(-i2\pi\omega/N)$.

Here is the formal proof of straight line phase/frequency.

$$\begin{aligned}
\hat{W}(\omega) &= \tfrac{1}{256}\textstyle\sum_{t=0}^{255} W_t N_{\omega t} \\
&= \tfrac{1}{256}[\textstyle\sum_{t=1}^{255} V_{t-1}N_{\omega t} + V_{255}N_{\omega 0}] \\
&= \tfrac{1}{256}[\textstyle\sum_{t=0}^{254} V_t N_{\omega(t+1)} + V_{255}N_{\omega 0}] \\
&= \tfrac{1}{256}\exp(\tfrac{-i2\pi\omega}{N})\textstyle\sum_{t=0}^{254} V_t N_{\omega t} = \exp(\tfrac{-i2\pi\omega}{N})\hat{V}(\omega)
\end{aligned}$$

Alternatively,

$$\hat{V}_\omega\overline{\hat{W}_\omega} = \hat{V}_\omega\overline{e^{\frac{-i2\pi\omega}{N}}\hat{V}_\omega} = e^{\frac{i2\pi\omega}{N}}|\hat{V}_\omega|^2. \tag{4.1}$$

Therefore, we can conclude that $\text{Phase}(\hat{W}) = 2\pi\omega/N + \text{Phase}(\hat{V})$.

4.4 Direction and Slope of Phase/Frequency Relationships

In the case of identical signals separated only by a lag, phase and frequency are perfectly correlated, the slope of the line depends only on the size of the time lag and the direction of the line depends only on which channel's signal lags behind that of the other. Figures 4.1 to 4.3 show the phase/frequency relationships for pairs of series where the components of each pair are identical except that one lags behind the other by a given fraction of an epoch. These lags were simulated using channel 29 of Dot Localization 1's EEG and rotating the first epoch by a constant. Identical but lagged signals can be simulated by choosing any channel and any epoch of any subject's EEG and rotating the series by a constant with the same results being achieved. The steepness of the line increases as the size of the lag increases and the direction of the line depends only on the direction of the lag.

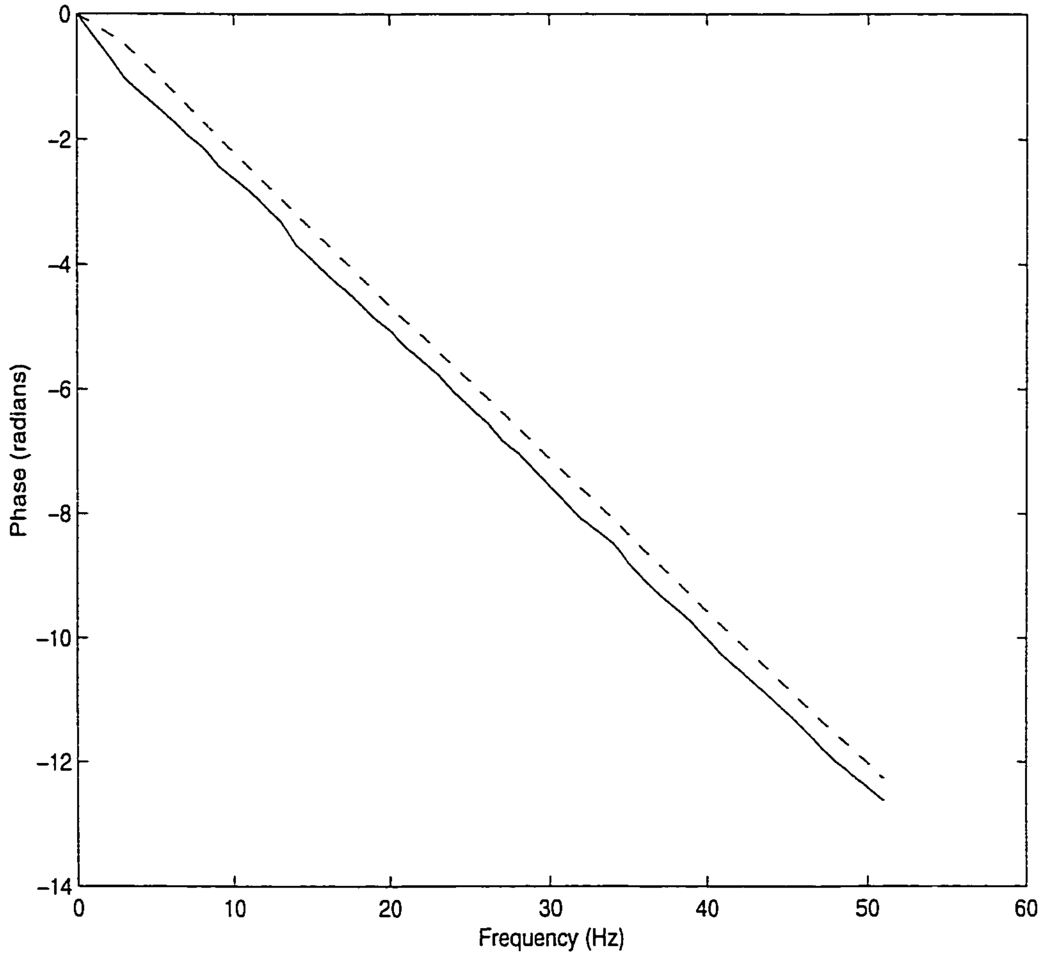

Figure 4.1: Phase/Frequency Relationship for a Perfect Lag of 10
The phase/frequency relationship for a pair of series where the first lags exactly 10/256 of an epoch behind the second results in a steep, downward sloping and almost perfectly straight line.

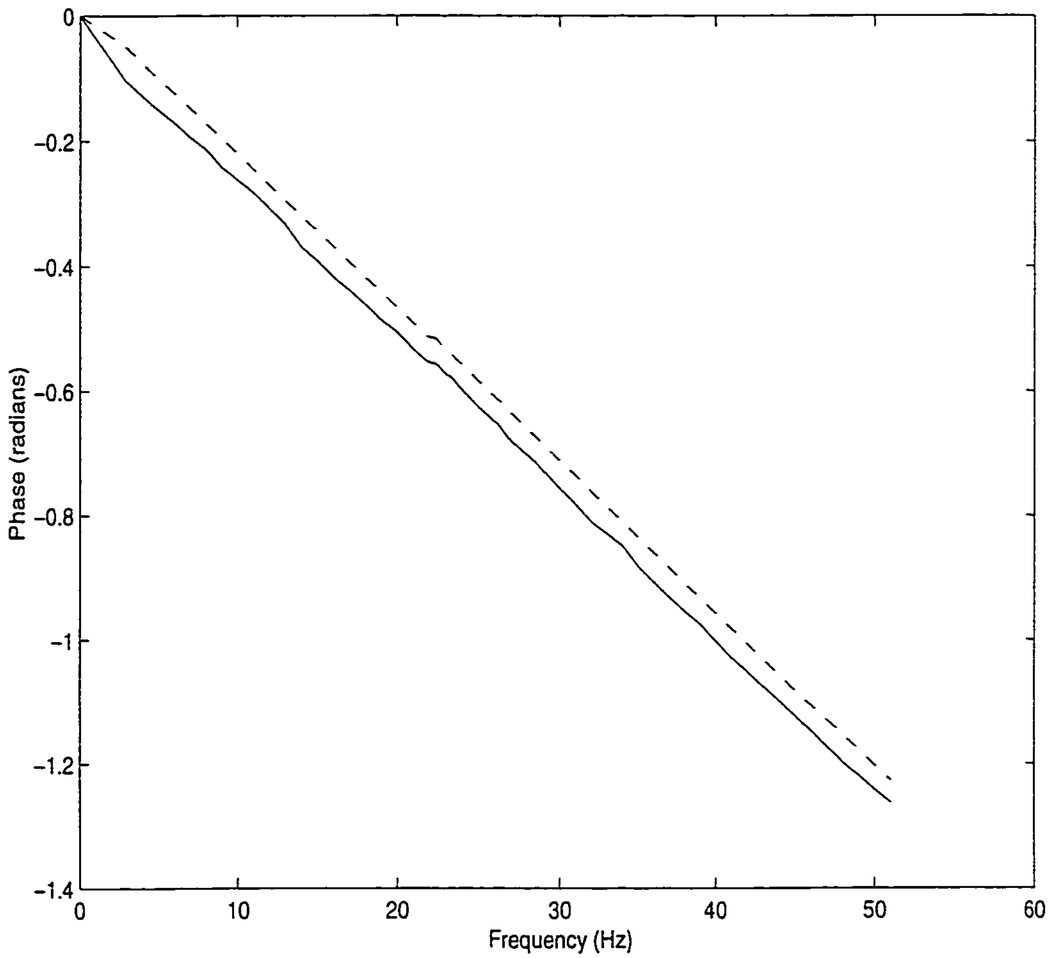

Figure 4.2: Phase/Frequency Relationship for a Perfect Lag of 1
This phase/frequency relationship occurs when the first series is a perfect lag of 1/256 of an epoch behind the second. It is still linear and downward sloping, but it is less steep (numerically) than the first.

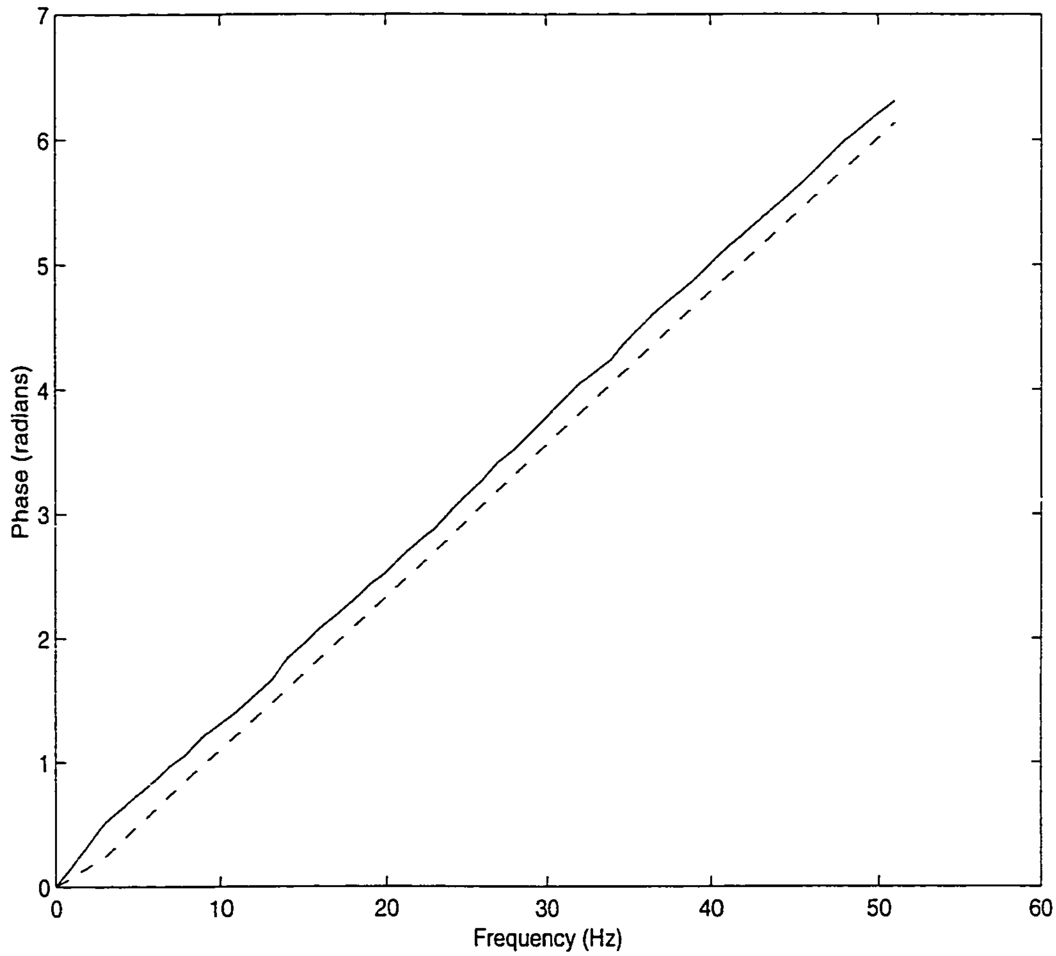

Figure 4.3: Phase/Frequency Relationship for a Perfect Lag of -5
The direction of the relationship between phase and frequency is reversed when the direction of the lag is reversed. Because the second series lags behind the first instead of the other way around, the relationship is positive instead of negative.

4.5 Potential Use of Channel Pair Phase/Frequency

If theory holds true, there will exist channel pairs in a typical subject's EEG data for which there are linear phase/frequency relationships. In the imperfect real world, the noise present in EEG signals carries over and obscures the linearity evident in theory. Nevertheless, hopes were that there would be channel pairs whose series were similar enough, had high enough lag and low enough noise for recognizable linear phase/frequency relationships to emerge from the EEG data of same-task subjects. It was also hoped that such relationships would differ enough among tasks to classify subjects.

Basic steps in finding such channel pairs involved computing channel pair covariances for all 43×43 channel pairs in each subject's EEG, storing the complex covariances as phase arrays and running a correlation program to find pairs where there was a high correlation between phase and frequency.

4.6 Trials of Phase/Frequency

Continuing to follow channel pair phase/frequency relationships as a route to discrimination required trials to see if it was computationally feasible and if linear relationships existed in the data.

Computing channel pair covariances for the 20 one-second epochs for each subject would give 20 chances for classification. The logical way to assess computational feasibility was to write and run Matlab programs to carry out the calculations and save the results because they were expensive. However, the four dimensional arrays which would have been produced for 88 subjects, each of size $43 \times 43 \times 256 \times 20$, were beyond computer capacity. As an alternative, covariances were computed across the between one-second time dimension, thus yielding $43 \times 43 \times 256$ arrays which could be stored. The smoothing of data in this process compensated for only having one chance instead of 20 to classify each subject according to task.

To see if linearity existed in practice, channel pair covariances were computed and the phases were stored as $43 \times 43 \times 256$ phase arrays, one for each subject. Dot Localization 1's phase array was then run through a correlation program. Linear regression analysis concluded that there was a significant phase/frequency correlation coefficient ($r = 0.9764$, $P < 0.01$) for at least one channel pair. This was sufficient enough proof of the existence of channel pairs yielding linear phase/frequency relationships to justify pursuing the methodology further.

4.7 Problems in Using Phase/Frequency

Processing data beyond the Fourier transform imposes one more procedure. Lead-lag theory requires analyzing the phase/frequency relationships for pairs of channels, not individual channels. Therefore, channel pair covariances must be computed, the phases of the covariances calculated and these stored as phase arrays.

Excessive noise obscures linearity and inhibits discriminant analysis. Noise reduction or smoothing techniques must be used and because some methods of reducing noise are more efficient than others, care is needed that EEG data is not wasted on inefficient smoothing techniques.

A third problem involves developing a specific discrimination methodology. The common classical approach to discrimination is the Maximum Likelihood Estimation Procedure (MLE) described in Morrison (1990) [10]. This procedure involves considering likelihood ratios, using a quadratic discrimination function and minimizing the expected cost of misclassification. Because MLE discrimination involves manipulating likelihoods and probability densities, it works best when there are just one or two parameters in the model. However, if MLE was to be applied to the channel pair phase/frequency case, the model would consist of either $43 \times 43 \times 256$ means and variances of phase or 43×43 slopes.

Without being able to apply elegant methodologies like MLE, specific conditions to classify subjects according to task must be hammered out. Moreover, the $43 \times 43 \times 256$ phase arrays are too large to use directly in discrimination. Data reduction is essential and reduced data must also contain information useful for classification. One approach is to use phase/frequency linearity to reduce the data in each subject's phase array to a channel pair which might be common to her task group. Another route is to reduce data to a channel pair for each task with phase/frequency linearity specific to that task.

Chapter 5

Construction of Phase Arrays

Transforming the frequency domain data to the phase data needed to obtain channel pair phase/frequency relationships follows these three steps.

The data is processed beyond the Fourier transform by computing the covariances for all 43×43 channel pairs and 256 frequencies across the between one-second time dimension for each subject. This gives a $43 \times 43 \times 256$ array of complex channel pair covariances for each subject. However, the covariances are very noisy.

To reduce the noise in channel pair covariances, data is smoothed over the complex number domain for the whole array of each subject.

The phases of the smoothed complex channel pair covariances are calculated and stored as phase arrays. These contain the data necessary to construct phase/frequency relationships for any channel pair for any subject.

5.1 Channel Pair Covariances

A Matlab program using loops and the covariance command was written to compute covariances for every channel pair and every frequency. These were computed across the between one-second time dimension, resulting in a total of $43\times43\times256$ covariances for each of the 88 subjects.

To describe this in formal terms, we denote V as the EEG data for one subject and set $\hat{V}$ to be its Fourier transform. Further, we let $\hat{V}_s$ $s = 1, \ldots, 20$ be the twenty one-second Fourier transformed data matrices. Then, for fixed channels a and b and a fixed frequency ω, we define the (sample) covariance by $C_{a,b,\omega} = \sum_{s=1}^{20} (\hat{V}_s)_{a\omega} \overline{(\hat{V}_s)_{b\omega}}$, where the bar denotes complex conjugate. Note that if $a = b$, then $C_{a,a,\omega} = \sum_{s=1}^{20} |\hat{V}_s|^2_{a\omega}$ is non-negative and real. We define C to be the $43 \times 43 \times 256$ array of all such covariances.

Since complex numbers are involved in calculating covariances, the resulting covariances are also complex and thus have phase. There is a different phase for every channel pair and for every frequency and so phase/frequency relationships can be computed for each channel pair for each subject. The exceptions are channel pairs of the same channel number like (1, 1) or (10, 10) where the covariances are real because the numbers in the two channels are the same.

Generally, computing covariances is not the same as computing the sum of products. Consider the following two series, x and y. If x denotes the numbers 1 to 10 and y the numbers 11 to 20, then $\sum_{k=1}^{10} x_k y_k = 935$ while $\mathrm{Cov}(x, y) = (\sum_{k=1}^{10} x_k y_k - (1/10) \sum_{k=1}^{10} x_k \sum_{k=1}^{10} y_k)/9 \simeq 9.16667$. However, in the case of using the phase of EEG data in discrimination, the two are essentially the same. First, because neither the mean nor the trend of any one-second epoch for any one channel contains any meaningful information, $C_{a,b,\omega} \approx C_{a,b,\omega} - (1/20) \sum_{s=1}^{20} (\hat{V}_s)_{a\omega} \overline{\sum_{s=1}^{20} (\hat{V}_s)_{b\omega}}$. Multiplying or dividing complex numbers by positive real constants only affects the magnitude and not the phase. Therefore, the same phase/frequency relationships will result whether product sums or covariances are computed.

5.2 Smoothing

The channel pair covariances can be very noisy and such noise should be reduced by smoothing techniques. Smoothing must be done over the complex number domain and phase calculated afterwards since the phase of an average is not equal to the average of the phases. Smoothing over the phase domain would result in the erroneous conclusion that phase does not depend on frequency. This is a misconception because phases are only congruent modulo 2π and so the mean phase would depend not on the nature of the EEG data but on the range used to compute the phase (0 to 2π or $-\pi$ to π).

Four methods of smoothing were employed to reduce noise in the channel pair covariances.

5.2.1 Epoch Aggregation

When covariances are computed across the between one-second time dimension (Section 5.1) instead of separately for each epoch (Section 4.6), data is smoothed by the averaging technique of epoch aggregation. Figures 5.1 and 5.2 show the phase/frequency relationships for channel pair (29, 42) for epochs one and two of Dot Localization 1's EEG data. Noise masks any similarity between the two plots. The phase/frequency relationship for channel pair (29, 42) for Dot Localization 1 after covariances are computed across the between one-second time dimension is displayed in Figure 5.3. The plot still exhibits too much noise to show a linear relationship, so further smoothing is required.

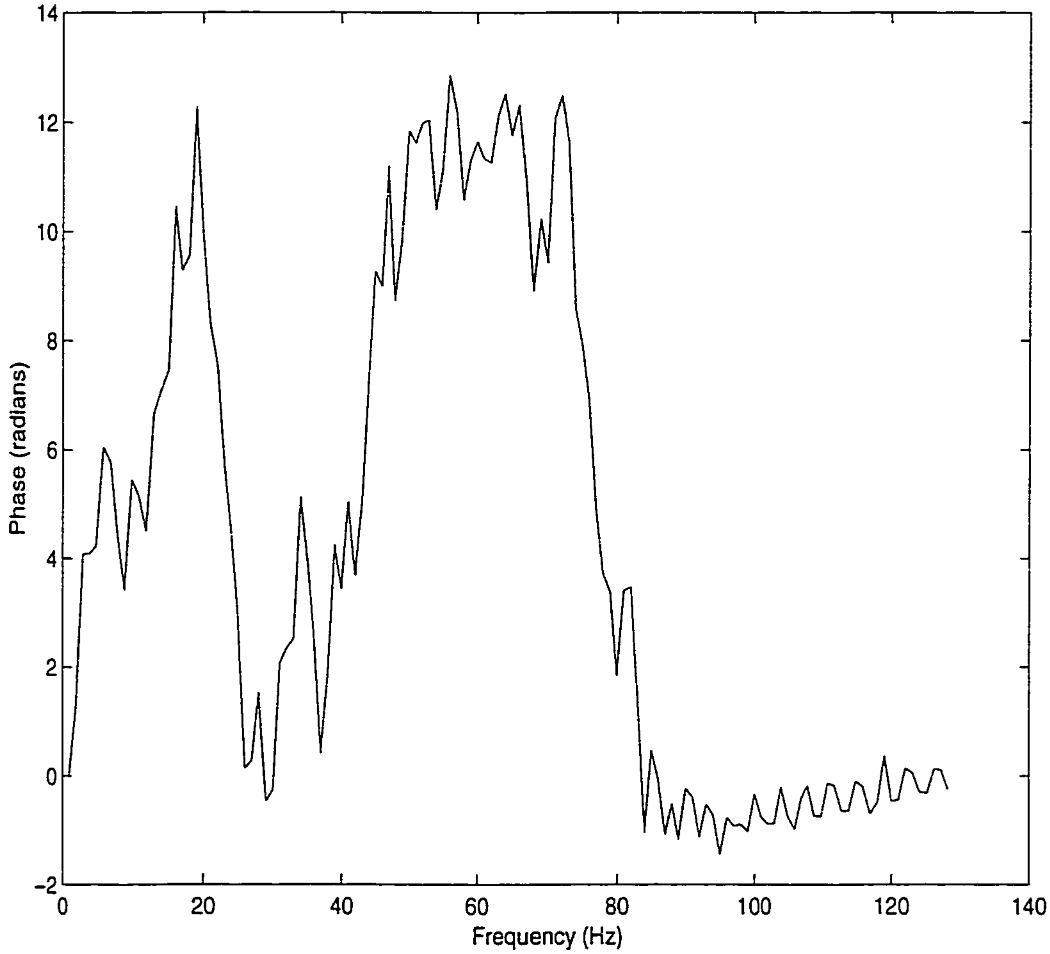

Figure 5.1: Phase/Frequency for Epoch One, No Epoch Aggregation
Without epoch aggregation being performed, the phase/frequency relationship for channel pair (29, 42) for the first epoch of Dot Localization 1's EEG contains too much noise to show a linear relationship.

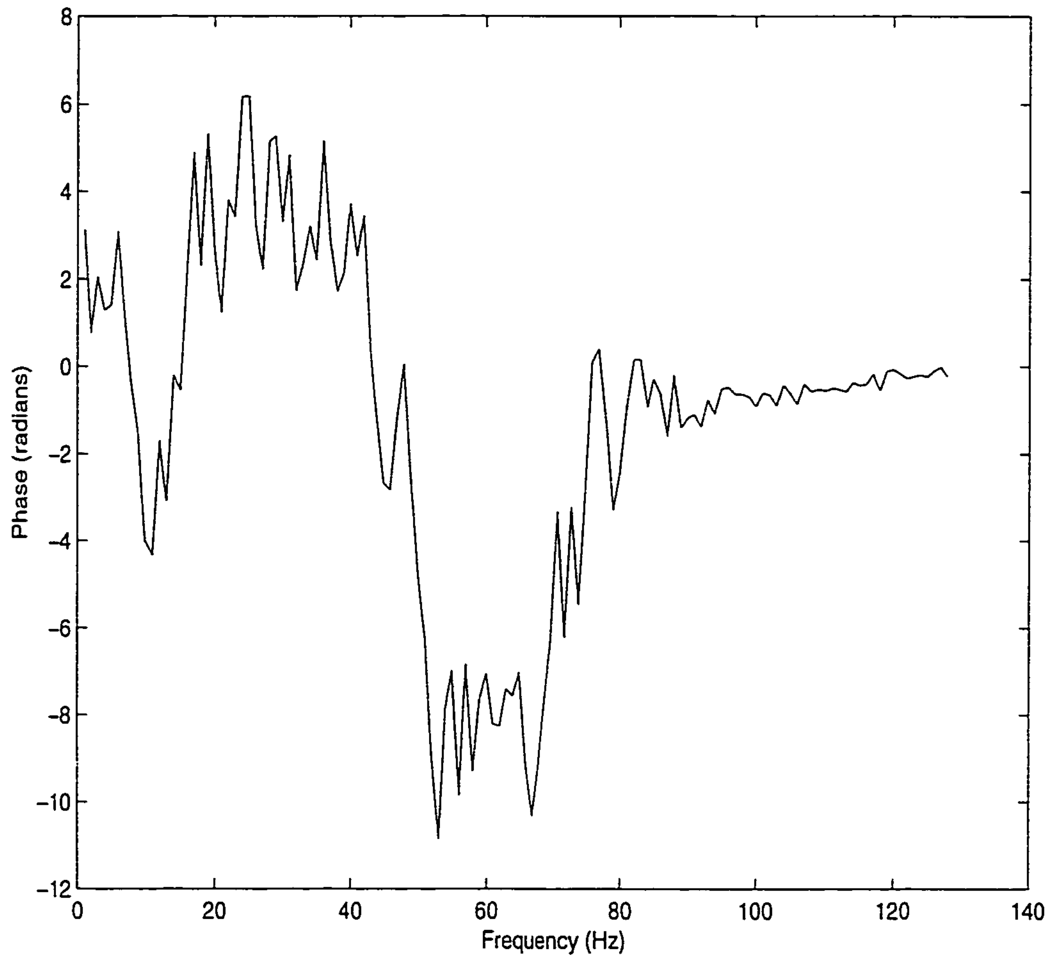

Figure 5.2: Phase/Frequency for Epoch Two, No Epoch Aggregation
The phase/frequency relationship for channel pair (29, 42) for the second epoch of Dot Localization 1's EEG is also very noisy. There is no similarity between this plot and the one above. Extreme noise causes phase/frequency relationships to behave very differently in different epochs.

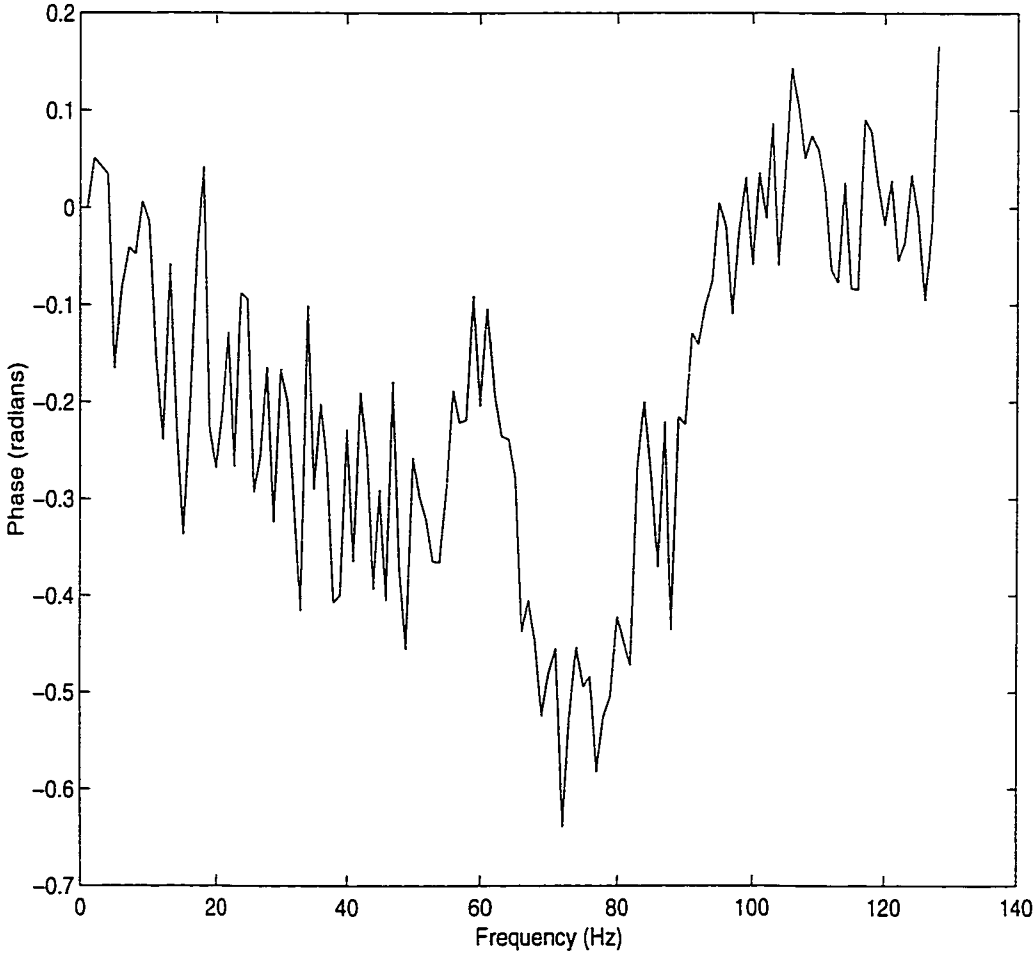

Figure 5.3: Phase/Frequency after Epoch Aggregation
With epoch aggregation but without any other smoothing, the phase/frequency relationship is still too noisy to show any linear relationship.

5.2.2 Moving Average

Smoothing was also accomplished by averaging across neighboring frequencies using a moving average of 5. Taking a moving average in the time domain is known to smooth time series (Shumway, 1988) [16]. Taking it in the frequency domain also smoothes by reducing large phase fluctuations between neighboring frequencies. However, it shortens the phase/frequency relationships, especially if the moving average is large, and therefore, it is not appropriate to use a moving average of more than 5. Moving average was implemented in Matlab as a separate function which looped through the 256 frequencies.

To write moving average formally, we let C be the pre-smoothed $43 \times 43 \times 256$ matrix of complex numbers, CMA be the $43 \times 43 \times 252$ matrix of complex numbers after a moving average of 5, a, b be channels and ω be frequency. Then, $CMA_{a,b,\omega} = (1/5)(C_{a,b,\omega} + C_{a,b,\omega+1} + C_{a,b,\omega+2} + C_{a,b,\omega+3} + C_{a,b,\omega+4})$ for $0 \leq \omega \leq 251$.

The phase/frequency relationship for channel pair (29, 42) for Dot Localization 1 after both epoch aggregation and moving average smoothing are applied is presented in Figure 5.4. There is less rapid fluctuation of phase with respect to frequency, but the presence of noise still masks any linear relationship.

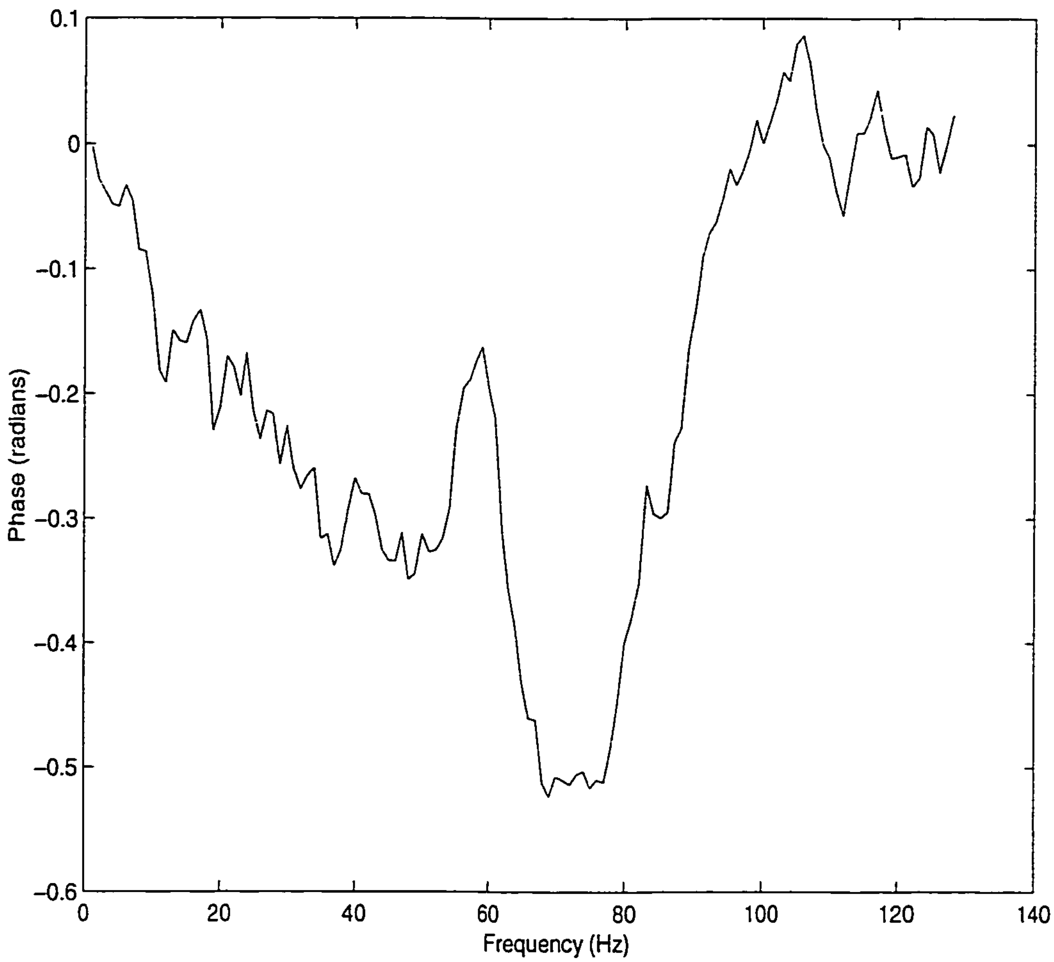

Figure 5.4: Phase/Frequency after Smoothing by Epoch Aggregation and Moving Average

Adding moving average further smoothes by reducing large phase fluctuations with respect to frequency.

5.2.3 Frequency Selection

Noise was also reduced by discarding frequencies with low signal to noise ratios. The EEG frequencies containing the highest signal to noise ratio are known to be those between 2 Hz and 50 Hz and so another smoothing technique is to consider only these frequencies. There is a theoretical cost. By discarding frequencies below 2 and above 50, the phase/frequency relationship is represented by a shorter series since only 49 instead of 252 frequencies are considered. However, it is a very successful smoothing technique and its application improves discrimination results.

In formal terms, CMA is the $43 \times 43 \times 252$ array of complex covariances, PMA is the array of corresponding phases, a, b are channel pairs and ω is frequency. Then $PMA = \text{phase}(CMA)$. Furthermore, for every pair (a, b), the phase/frequency relationship is Y versus X where $Y = PMA_{a,b,\omega}$ and $X = \omega$ where $2 \leq \omega \leq 50$ instead of $0 \leq \omega \leq 251$.

Figure 5.5 displays the phase/frequency relationship for Dot Localization 1, channel pair (29, 42) after epoch aggregation, moving average and frequency selection smoothing. We see that the relationship between phase and frequency is clearly negative because noise is reduced enough to show a linear trend.

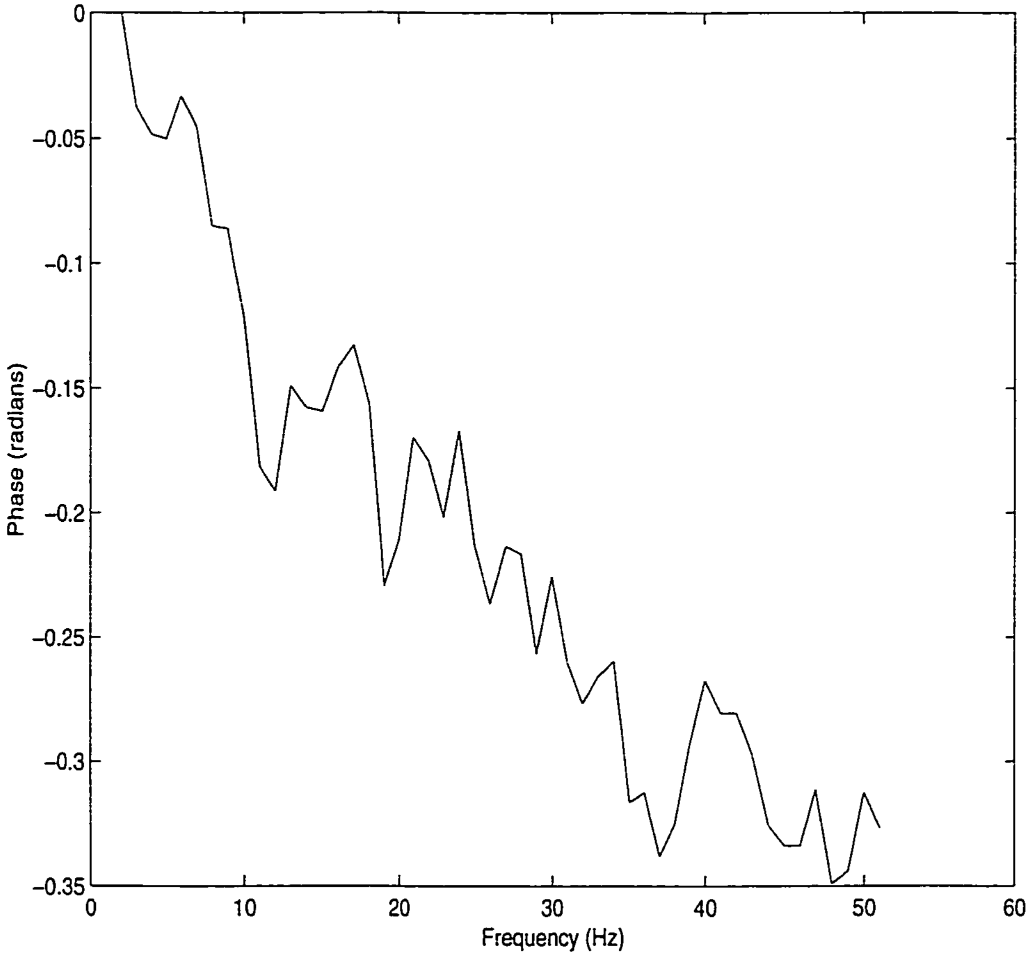

Figure 5.5: Phase/Frequency after Smoothing by Epoch Aggregation, Moving Average and Frequency Selection

Adding frequency selection to the smoothing process removes the noise of the high frequencies so that linear relationships become apparent.

5.2.4 Principal Component Analysis

Principal component analysis (PCA) was performed to improve the conditioning of the matrices. Applying it to a 43×43 covariance matrix is beneficial for the same reasons that it is good for detecting multiple collinearity in regression (Rawlings, 1988) [14]. If two column vectors or row vectors are close to but not quite parallel, it inflates noise in phase/frequency relationships.

Every 43×43 matrix of covariances is conjugate symmetric. That is, if T is a matrix of covariances, then T is its own conjugate transpose ($T = T'$). Therefore, by the spectral decomposition theorem, it follows that T is orthonormally diagonalizable, that is, there exist matrices U and D such that $T = UDU'$. Furthermore, U is orthonormal ($UU' = U'U = I$ where I is the identity matrix) and D is diagonal.

More specifically, U is the matrix of eigenvectors of T and the diagonal elements of D are the corresponding eigenvalues. If D consists of eigenvalues which are small compared with other eigenvalues, then the matrix T is ill-conditioned and this causes the data to be noisy. The problem is solved by setting small eigenvalues to 0 so that D becomes D_R. After this, $T_R = UD_RU'$ is computed. The cost of setting eigenvalues to 0 is the danger of reducing signal as well as noise. Therefore, D_R is computed in such a way that trace(abs(D_R)) is approximately 0.9trace(abs(D)). In Matlab, this was accomplished by using the singular value decomposition command. Because the eigenvalues of D were in descending order of absolute value, it was sufficient to loop backward through the diagonal elements of D to set the small eigenvalues to 0.

If a is a scalar, it is an eigenvalue of T if and only if $\det(T - aI) = 0$. A parallel concept is that if x is a vector, it is a corresponding eigenvector to eigenvalue a if and only if $Vx = ax$. If x_1 and x_2 are different column vectors of U, then it follows that: x_1 and x_2 are both eigenvectors of T; $x_1.x_2 = 0$, that is, x's are orthogonal; and $x_1.x_1 = x_2.x_2 = 1$, that is, x's are of length 1 and hence orthonormal. Furthermore, the set containing all the columns of U form a basis for the eigenspace of T. In the case of applying PCA to EEG-based discrimination, T is the 43×43 matrix of covariances computed for just one frequency ω.

For every frequency ω, T is set to $CMA_{a=1:43,b=1:43}$ to obtain the 43×43 matrix before PCA. T_R denotes the matrix obtained by applying PCA. That is, if $T = UDU'$, then $T_R = UD_RU'$.

Repeating this procedure for every frequency ω turns the array of complex covariances before PCA, CMA, into the array of complex covariances after PCA, CPD. The procedure was repeated for all frequencies, but due to frequency 2 to 50 selection, the same results could have come from performing PCA for only frequencies $2 \leq \omega \leq 50$.

Figure 5.6 shows the phase/frequency relationship for channel pair (29, 42) for Dot Localization 1 after epoch aggregation, moving average, frequency selection and PCA smoothing. PCA has little effect on reducing the level of noise.

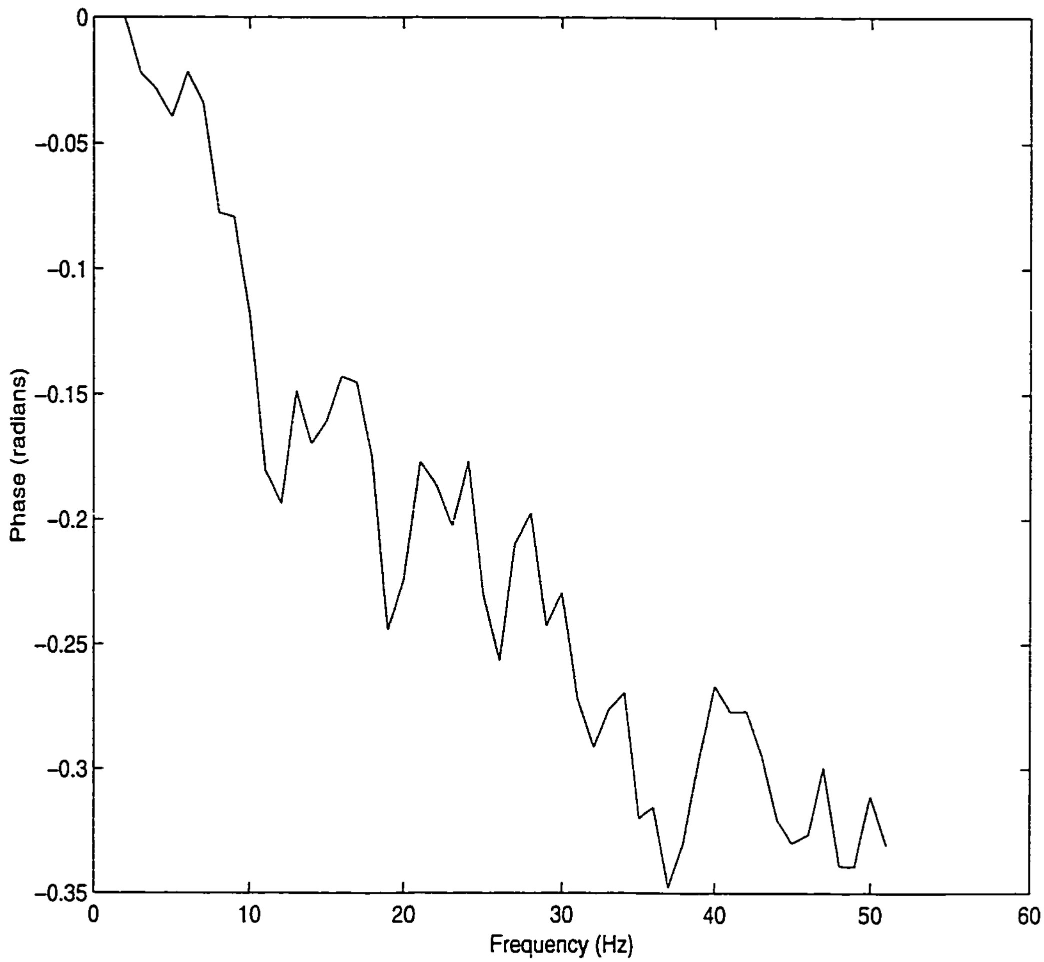

Figure 5.6: Phase/Frequency after Smoothing by Epoch Aggregation, Moving Average, Frequency Selection and PCA
The phase/frequency relationship after the addition of PCA to smoothing is nearly the same as it was before. PCA is the least effective smoothing technique.

5.2.5 Comparative Roles of Smoothing Techniques

Finally, to illustrate the specific roles of the smoothing techniques in 5.2.2, 5.2.3 and 5.2.4, it is instructive to look at Figures 5.7 to 5.9 which show the results of selective smoothing for Dot Localization 1, channel pair (29, 42). Figure 5.7 shows the rapid phase fluctuation with respect to frequency when frequency selection and PCA are performed, but when moving average is not. Whenever moving average smoothing is not performed, phase fluctuates rapidly with respect to frequency. The apparent non-linear relationship when moving average and PCA are applied, but when frequency selection is not, is displayed in Figure 5.8. Whenever frequency selection is not performed, the noise beyond 50 Hz gives the illusion of non-linear relationships. Figure 5.9 illustrates how PCA by itself is not enough smoothing to reduce noise to a reasonable level. Comparing moving average, frequency selection and PCA, the latter is the least effective smoothing technique.

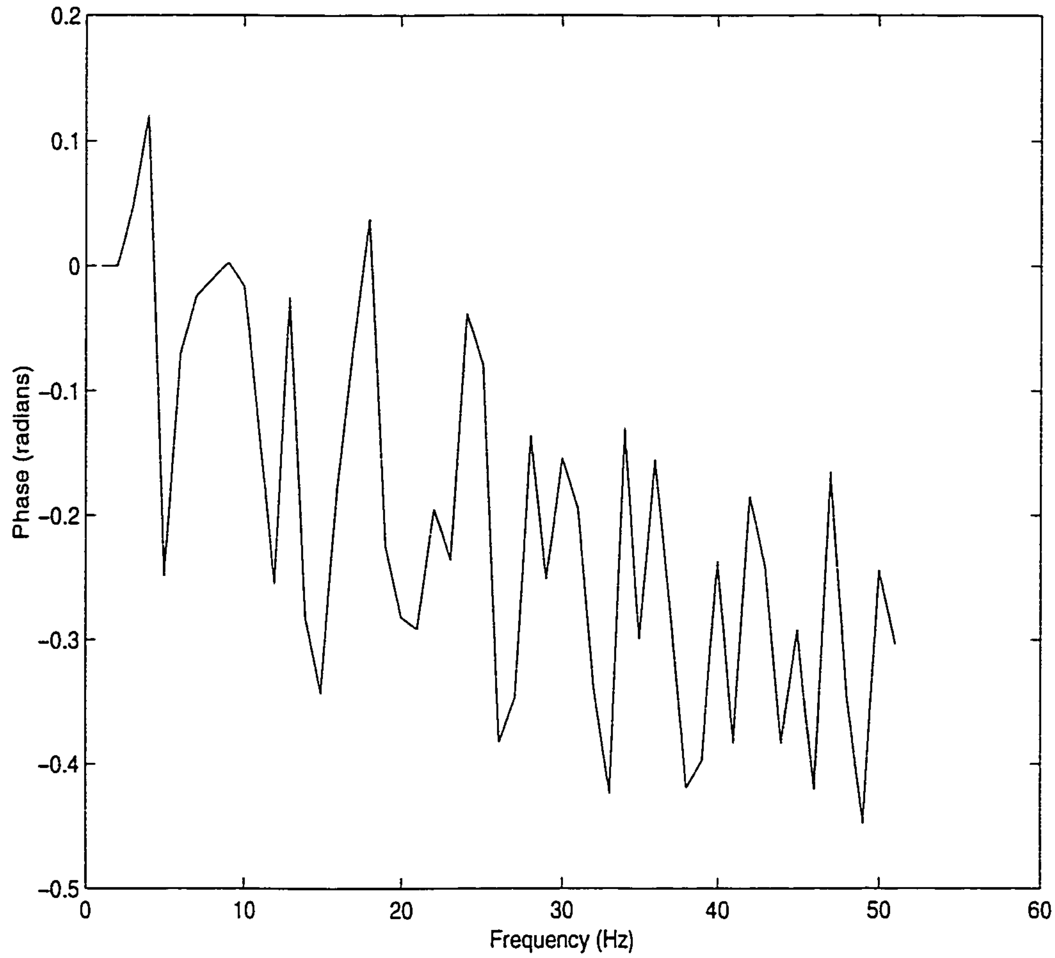

Figure 5.7: Phase/Frequency Smoothing with Frequency Selection and PCA but without Moving Average

When moving average is not performed, there are rapid fluctuations in phase with respect to frequency. Moving average helps by removing these fluctuations.

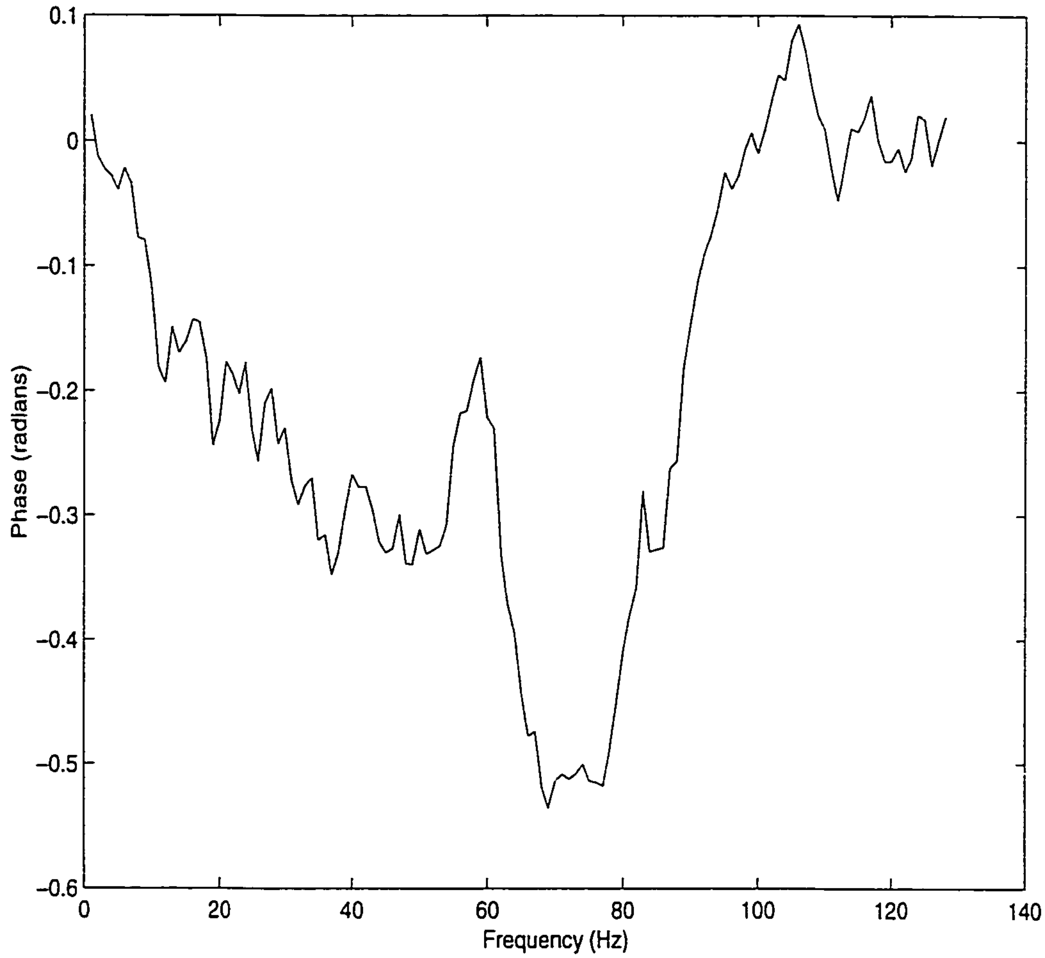

Figure 5.8: Phase/Frequency Smoothing with PCA and Moving Average but without Frequency Selection

Not performing frequency selection produces the illusion of non-linearity in phase/frequency plots. This is because anything beyond 50 Hz in EEGs is highly noise.

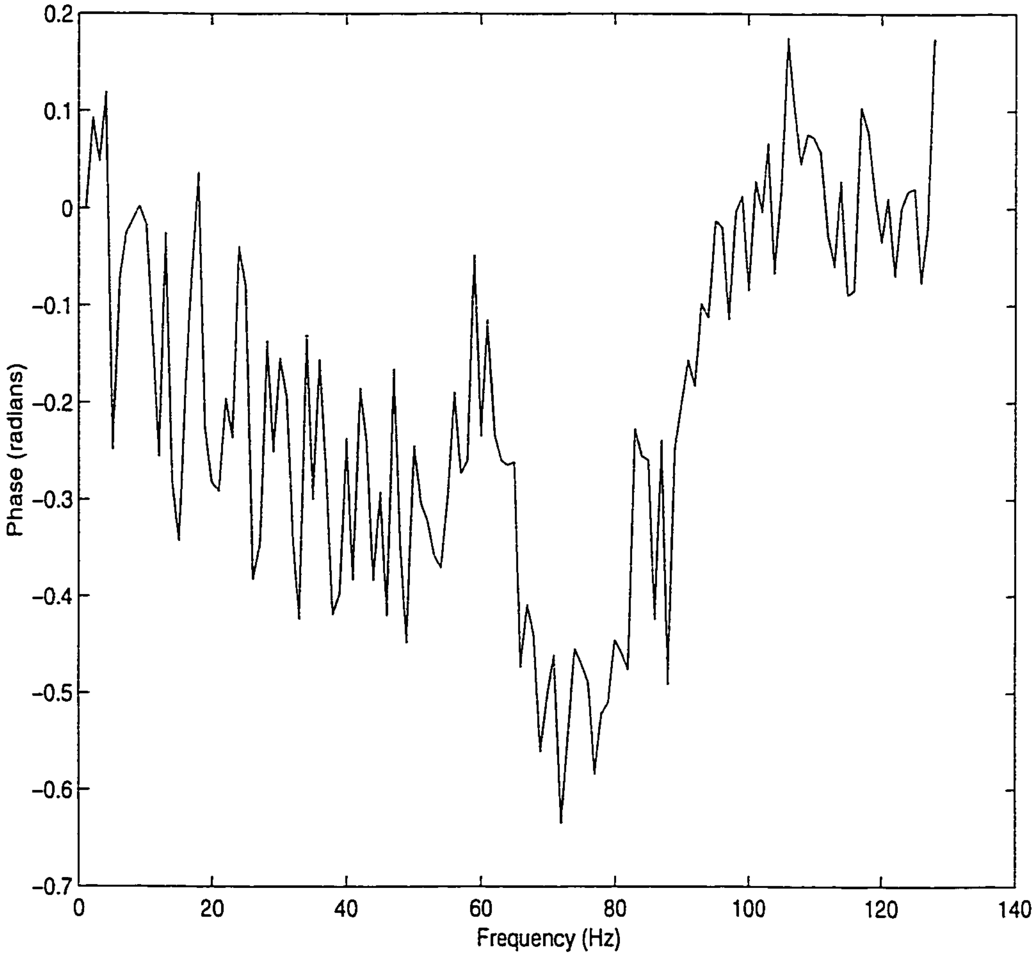

Figure 5.9: Phase/Frequency Smoothing with only PCA
PCA without moving average or frequency selection is ineffective. Neither the rapid phase fluctuations nor the high frequency noise is removed.

5.3 Phase Arrays

Once channel pair covariances are computed and smoothing is completed, the Matlab angle command can be used to calculate the phases of the complex covariances. These are then stored as $43 \times 43 \times 252$ phase arrays for every subject. This is the same as having a set of 43×43 phase/frequency relationships for each. Any phase/frequency relationship can be viewed by taking the phase series for the channel pair under consideration and putting it with the frequency series $2 \leq \omega \leq 50$.

Chapter 6

Use of Phase Arrays

It would be impractical to use the huge phase arrays and all the phase/frequency relationships they embody to discriminate between tasks. Statistical procedures must be used to reduce data to a form where channel pairs with potential task specific linearity can be found. Task dependent characteristics must then be translated into reliable and efficient rules to classify each subject according to her assigned task. The rules must then be tested.

This chapter looks at strategies to reduce data, discusses matters to consider when developing rules, reviews the principles of efficiency and reliability and defines the rule development and test data sets.

6.1 Data Reduction

Data reduction must be done so as to find channel pairs with phase/frequency linearity which might differ between tasks. Linearity can be measured by correlation, which emphasizes signal similarity, and slope, which emphasizes signal lag.

Measures of linearity can be applied to each subject or to each task. When applied to subjects, the goal is to find channel pairs which are common to same-task subjects.

When applied to tasks, the aim is to find a channel pair for each task with high phase/frequency linearity specific to that task.

Finding these channel pairs is somewhat of a trial and error exercise.

6.1.1 Subject Channel Pairs

To find subject channel pairs, these techniques were tried:

(i) Finding the channel pair for each subject with highest least squares phase/frequency slope;

(ii) Determining the channel pairs for each subject with phase/frequency correlations above 0.95;

(iii) Computing the channel pair for each subject with highest correlation between phase and frequency.

The procedures for these trials and their results are reported in Chapter 7.

6.1.2 Task Channel Pairs

These are the techniques used to find task channel pairs:

(i) Computing the channel pair with highest correlation between phase and frequency for each task;

(ii) Determining the channel pair with highest least squares phase/frequency slope for each task.

Finding the channel pairs is only half the procedure. Their phase/frequency properties must be examined for differences between tasks. This can be done by:

(i) Plotting the phase/frequency relationships and visually examining them to produce categorical differences between tasks;

(ii) Using correlation/slope vectors to obtain numerical boundaries between tasks.

The procedures for finding task channel pairs and their categorical or numerical phase/frequency properties, along with the results of their trials, are reported in Chapters 8, 9 and 10.

6.2 Rules for Classifying Subjects According to Task: Direct Identification Rules and Ruling Out Conditions

Rules for classifying subjects according to their assigned task can be developed from the subject channel pairs and from the categorical or numerical phase/frequency properties of the task channel pairs.

Easy to use rules would be able to identify each subject's task directly. For example, if it was found that the highest correlation subject channel pair for all dot localization subjects was (10,13) as opposed to (20,23) for all word finding, (30,33) for all eyes open and (40,43) for all eyes closed, then task identification would be simple and direct.

In reality, the complexities of within-task variation or overlapping between tasks may make it impossible to directly classify subjects. For example, if channel pair (10,13) occurred for all dot localization subjects, half of all word finding subjects and not at all for eyes open and eyes closed subjects, then none of the four tasks could be identified directly. The best that could be done would be to develop ruling out conditions. So the presence of channel pair (10,13) would rule out eyes open and eyes closed and the absence of (10,13) would rule out dot localization.

6.3 Reliabilities and Efficiencies of Rules

A good direct task identification rule or ruling out condition should satisfy high reliability and moderate to high efficiency standards. It seemed reasonable to aim for 0.8 reliability and 0.5 efficiency.

Reliability is a measure of how often a classification rule makes a Type I error, that is, makes a wrong decision about a subject's task. The more often a direct rule misidentifies the task or the more often a ruling out condition erroneously eliminates the task, the lower the reliability. For example, a condition developed to rule out dot localization that rules out 10% of subjects actually performing dot localization has a reliability of 0.9.

The efficiency of a classification rule is a measure of how often a direct rule correctly identifies a task or how often a ruling out condition correctly rules out a task. Failing to directly identify a subject's task or failing to rule out a task that the subject is not performing is a Type II error. For example, a condition developed to rule out dot localization that actually rules out 70% of subjects not performing dot localization has an efficiency of 0.7.

If the sum of the reliability and efficiency of a direct rule falls below 1, it means that it misidentifies the task more often than it correctly identifies it. If a ruling out condition's reliability and efficiency sums to less than 1, it means that it erroneously rules out a task more often than it correctly rules it out.

A Type I error is more serious than a Type II error and this makes reliability more important than efficiency. Making a wrong decision about a subject's task or analogously, a wrong decision about a patient's psychiatric state, either by misidentifying it or erroneously ruling it out, is worse than making no decision at all.

6.4 Rule Development and Test Data Sets

So that rules could be tested, randomly ordered sets of the four task groups were split in two. The first set of 45 subjects (11 dot localization, 10 word finding, 12 eyes open and 12 eyes closed) was used to develop classification rules, while the remaining 43 were used to test the rules. The sets are unavoidably small.

Chapter 7

Categorical Classification Using Subject Channel Pairs

The task that a control subject performs should highly determine which pairs of channels receive similar but lagged signals. So theoretically, there should be certain pairs of channels with high phase/frequency correlation and/or high phase/frequency slope common to subjects in the same task group but different from those of other task groups.

It was thought that such distinctive channel pairs might emerge from each subject's phase array by obtaining the pair with highest least squares slope between phase and frequency, or by determining all pairs with phase/frequency correlation above 0.95, or by finding the pair with highest correlation between phase and frequency.

Only the last one succeeded in producing some task dependent patterns.

To review data processing to this point, channel pair covariances have been computed to give a $43 \times 43 \times 256$ array for each of the 88 subjects (Section 5.1). The covariances have been smoothed (Section 5.2). Phases of the smoothed complex covariances have been calculated to produce individual $43 \times 43 \times 252$ phase arrays and these contain all 43×43 channel pair phase/frequency relationships (Section 5.3).

7.1 Subject Channel Pairs with Highest Least Squares Slopes

Correlation versus least squares slope scatterplots for all 43 × 43 channel pairs for selected subjects reveal few high slopes, suggesting the possibility that they might be useful in discrimination. Figure 7.1 is an example.

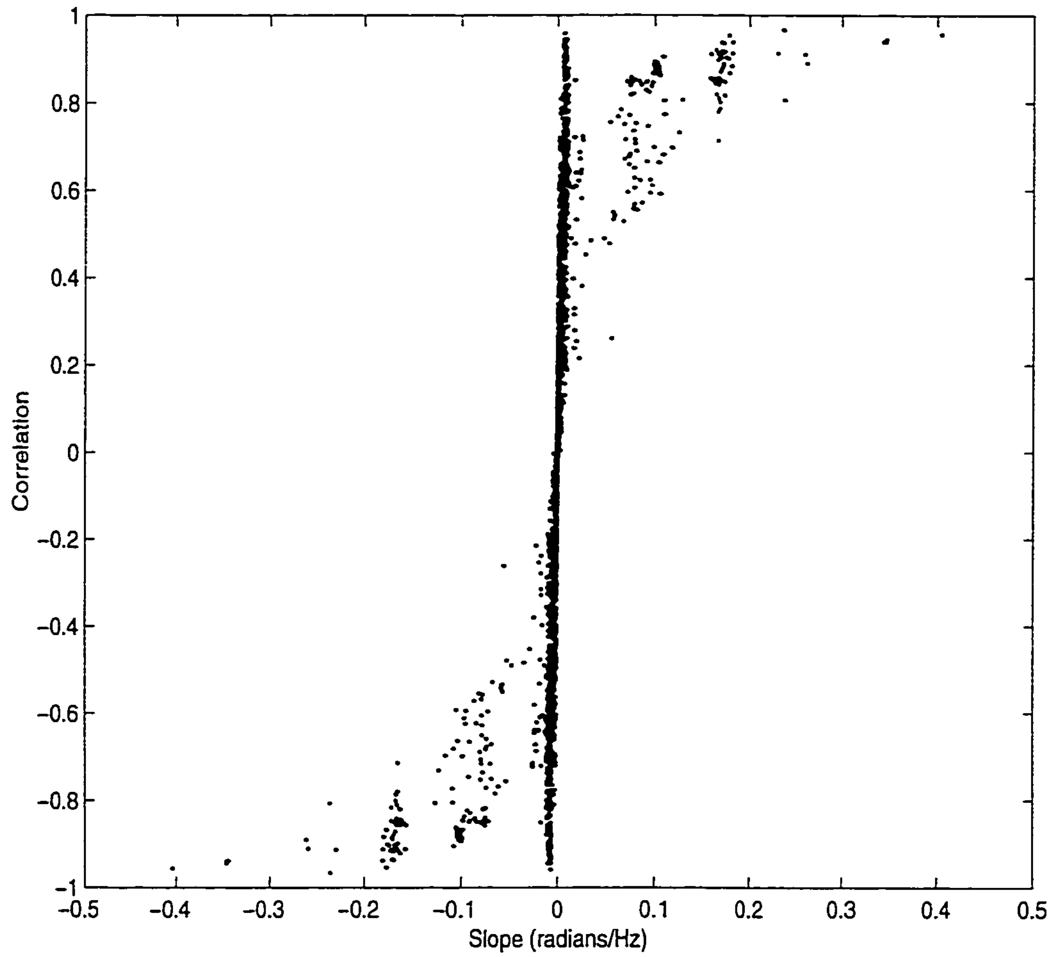

Figure 7.1: Correlation/Least Squares Slope Scatterplot, Dot Localization 1
Since there were few high slopes, their potential in discrimination was investigated.

Each subject's phase array was transformed to a 43 × 43 least squares slopes array and the channel pair with highest least squares slope in absolute value was found. As Table 7.1 shows, there are no task dependent patterns in the channel numbers for the rule data set. For example, most of the channel pairs tend to have numbers in the 30's or 40's, no matter what the task.

Table 7.1: Highest Least Squares Slope Subject Channel Pairs and Values, Rule Data Set

Dot Localization	Word Finding	Eyes Open	Eyes Closed
(11, 31) 0.4041	(32, 41) -0.2760	(32, 41) -0.3293	(1, 43) -0.2330
(11, 31) -0.3332	(32, 41) -0.3094	(37, 42) -0.2759	(35, 40) 0.3024
(14, 30) -0.3768	(22, 32) 0.2756	(2, 42) -0.3262	(3, 42) -0.2881
(36, 37) -0.2845	(13, 33) 0.2874	(36, 37) -0.2577	(1, 12) 0.2490
(31, 42) -0.0274	(32, 43) -0.0972	(31, 42) -0.1527	(33, 40) 0.1384
(35, 36) -0.2130	(2, 29) 0.3599	(24, 42) -0.1775	(11, 31) -0.1814
(4, 32) -0.1879	(12, 36) -0.2124	(2, 40) -0.1426	(11, 33) -0.3318
(32, 39) -0.3542	(11, 33) 0.2498	(29, 31) 0.3318	(2, 40) 0.3025
(3, 39) 0.1967	(11, 39) -0.1804	(14, 38) 0.2882	(39, 40) 0.4839
(2, 40) -0.3240	(11, 31) 0.2518	(12, 38) -0.4368	(31, 40) 0.3892
(13, 33) -0.1794		(24, 40) -0.1944	(1, 42) 0.3150
		(32, 41) 0.2322	(11, 35) 0.1587

7.2 Subject Channel Pairs with Correlations above 0.95

Histograms of correlation coefficients between phase and frequency for individual phase arrays, like the one in Figure 7.2, show that high correlations are comparatively few. This suggests that channel pair correlations above a certain level might have potential discrimination value.

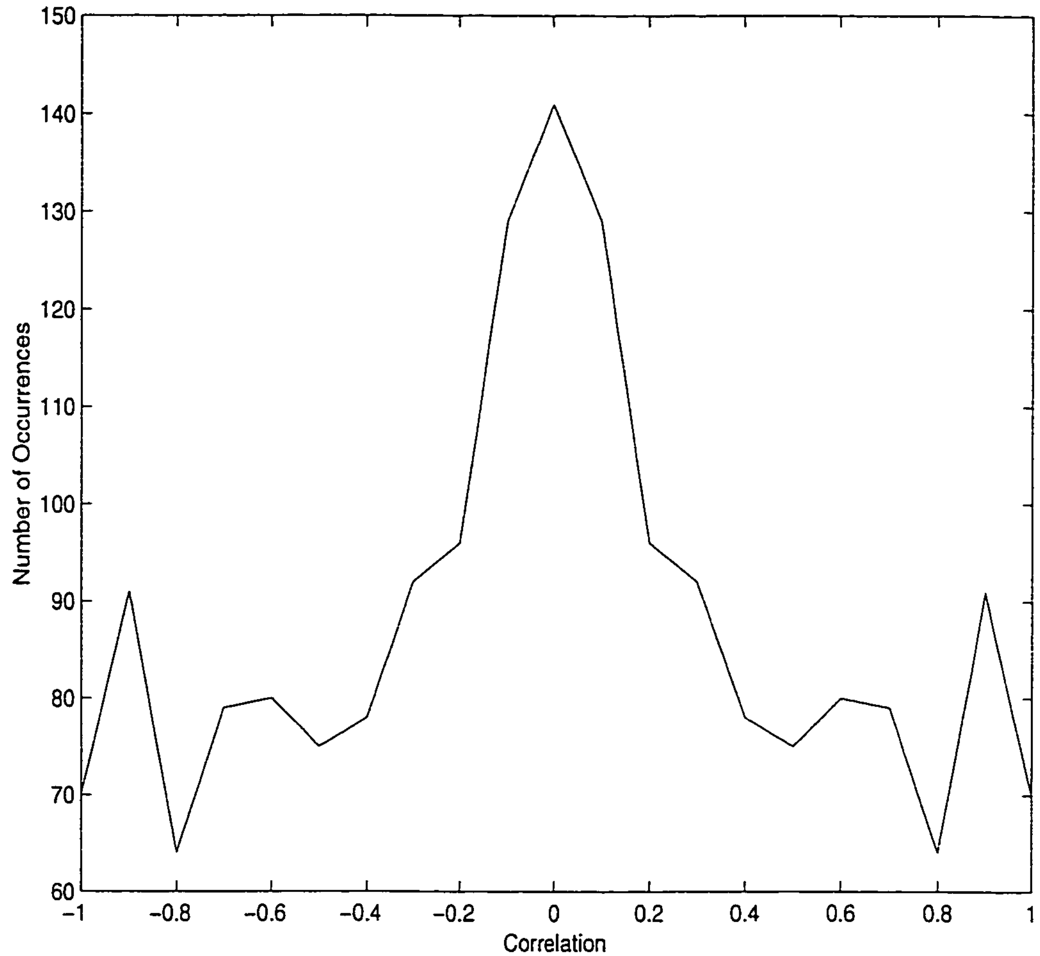

Figure 7.2: Histogram of Correlation Coefficients, Dot Localization 1
The few correlation coefficients at the high ends were examined for discrimination possibilities.

An attempt was made to list the channel pairs with correlations above 0.95 for all phase arrays. The procedure was abandoned when it was found that there were subjects with literally scores of such channel pairs.

7.3 Subject Channel Pairs with Highest Correlation

With the histograms of correlation coefficients in mind, every subject's phase array was processed to determine the channel pair which produced the highest correlation coefficient in absolute value between phase and frequency.

7.3.1 Procedure to Find Highest Correlation Subject Channel Pairs

Processing starts with CPD which is the matrix of complex covariances after all smoothing. Then the phase of the covariances are computed to get $PPD = \text{phase}(CPD)$. Channels a, b ($a \neq b$) are determined so that if $Y = PPD_{a,b,\omega=2:50}$ and $X = \omega$ where $2 \leq \omega \leq 50$, then $|\text{correl}(Y, X)|$ is a maximum. More formally, for all other channels c, d: we set $X = \omega$ where $2 \leq \omega \leq 50$; $Y_1 = PPD_{a,b,\omega=2:50}$; and $Y_2 = PPD_{c,d,\omega=2:50}$. We then compute $r_1 = \text{correl}(Y_1, X)$ and $r_2 = \text{correl}(Y_2, X)$. Finally, by definition, we know that (a, b) maximizes the correlation coefficient if and only if $|r_2| \leq |r_1|$ for every (c, d).

To perform this procedure in Matlab, a correlation program using the absolute value command and the maximum command was run on the phase array. It was then run again on the same phase array and stopped when it found the channel pair yielding the highest absolute correlation coefficient.

Note that because the length of the series is always 49, the F test statistic for non-zero slope depends only on $|r|$. The proof involves a string of equalities and a justification for each one:

$$F_{1,47} = MSR/MSE = SSR/(SSE/47) = 47SSR/SSE = 47r^2/(1-r^2).$$

The first equality follows from standard linear regression principles. The second is valid since it only involves writing MSR and MSE in terms of SSR and SSE. The third step follows from multiplying both the top and bottom by 47. The fourth manipulation involves dividing the top and bottom by SST.

Since squaring r eliminates the sign, F depends only on $|r|$.

7.3.2 Results of Highest Correlation Subject Channel Pairs Method

The highest correlation subject channel pairs technique reduces data to 88 channel pairs. The pairs for the rule data set are on Table 7.2. Although it would have been unreasonable to expect that the channel pair yielding the highest correlation would be the same for every same-task subject, the amount of overlapping of channel pairs among tasks was unexpected. Nevertheless, classification rules could be developed.

Table 7.2: Highest Correlation Subject Channel Pairs and Coefficients, Rule Data Set

Dot Localization	Word Finding	Eyes Open	Eyes Closed
(2, 31) -0.9665	(22, 43) -0.9589	(31, 42) +0.9551	(2, 33) -0.9357
(29, 42) -0.9832	(5, 32) -0.9620	(5, 34) -0.9450	(2, 35) -0.9876
(12, 41) -0.9760	(22, 41) -0.9747	(31, 42) +0.9429	(1, 30) -0.9388
(22, 43) -0.9852	(33, 40) -0.9643	(2, 35) -0.9825	(2, 35) -0.9657
(24, 35) -0.9845	(2, 35) -0.9728	(6, 35) -0.9732	(6, 35) -0.9765
(22, 43) -0.9638	(22, 43) -0.9787	(4, 33) -0.9720	(2, 35) -0.9643
(17, 40) -0.9744	(4, 37) -0.9729	(3, 36) -0.9811	(2, 31) -0.9800
(11, 40) -0.9536	(11, 33) +0.9759	(12, 32) +0.9657	(11, 40) -0.9702
(22, 43) -0.9629	(5, 34) -0.9677	(5, 34) -0.9580	(11, 33) -0.9516
(22, 43) -0.9737	(5, 38) +0.9498	(2, 35) -0.9736	(37, 40) +0.9836
(5, 34) -0.9600		(2, 35) -0.9587	(22, 43) -0.9679
		(2, 31) -0.9713	(13, 40) -0.9621

7.3.3 Ruling Out Conditions

Rules to classify subjects by directly identifying tasks are not to be found because word finding channel pairs behave randomly. The best that can be done is to find conditions which rule out dot localization, eyes open and eyes closed.

- Rule out dot localization if one channel pair number is less than 10.

- Rule out eyes open if one channel pair number is in the 20's.

- Rule out eyes closed if one channel pair number is in the 20's.

The presence of a number in the 20's separates active tasks from passive ones because channel numbers in the two passives overlap.

Table 7.3 gives the reliabilities and efficiencies of these conditions. The efficiencies of eyes open and eyes closed are far below 0.5, but reliabilities are good.

Table 7.3: Reliabilities and Efficiencies, Rule Data Set

Task	Reliability	Efficiency
Dot Localization	9/11	21/34
Word Finding	-	-
Eyes Open	1	10/33
Eyes Closed	11/12	9/33

7.3.4 Test of Ruling Out Conditions

The highest correlation subject channel pairs for the test data set are listed on Table 7.4. The ruling out conditions were applied to this set and their reliabilities and efficiencies calculated (Table 7.5).

The three ruling out conditions pass their tests since reliabilities and efficiencies sum to over 1, meaning that the probability of a Type I error does not exceed the probability of a correct classification. However, the reliability for dot localization drops sharply and the efficiencies for eyes open and eyes closed are low.

Table 7.4: Highest Correlation Subject Channel Pairs and Coefficients, Test Data Set

Dot Localization	Word Finding	Eyes Open	Eyes Closed
(13, 31) -0.9480	(22, 41) -0.9489	(11, 33) -0.9328	(37, 40) -0.9455
(22, 41) -0.9358	(6, 35) -0.9644	(12, 39) -0.9240	(24, 34) -0.9081
(23, 35) -0.9680	(10, 26) -0.9721	(24, 35) -0.9688	(24, 35) -0.9770
(9, 26) -0.9647	(9, 26) -0.9885	(9, 15) -0.9496	(14, 39) -0.9623
(9, 26) -0.9798	(9, 26) -0.9687	(12, 39) -0.9613	(14, 18) -0.9690
(9, 15) -0.9583	(9, 25) -0.9911	(9, 15) -0.9490	(9, 26) -0.9643
(10, 15) -0.9829	(9, 26) -0.9842	(9, 15) -0.9771	(9, 15) -0.9762
(9, 15) -0.9888	(6, 25) -0.9886	(9, 26) -0.9555	(12, 39) -0.9494
(9, 26) -0.9912	(9, 26) -0.9901	(9, 26) -0.9670	(37, 40) +0.9768
(9, 15) -0.9973	(10, 26) -0.9803	(10, 26) -0.9892	(10, 26) -0.9791
(10, 26) -0.9824		(8, 16) -0.9600	(36, 43) +0.9609

Table 7.5: Reliabilities and Efficiencies, Test Data Set

Task	Reliability	Efficiency
Dot Localization	3/11	26/32
Word Finding	_	_
Eyes Open	8/11	12/32
Eyes Closed	9/11	13/32

7.4 Discussion

Highest correlation subject channel pairs are the only ones resulting in rules. However, they do not allow direct task identification and can only provide ruling out conditions for three tasks. Although passed, test results are poor, with low efficiencies and decreases in reliability.

Problems arise from the random behavior of word finding channel pairs and channel number overlapping between tasks.

Chapter 8

Categorical Classification Using Highest Correlation Task Channel Pairs

The goal of using highest correlation task channel pairs is to reduce data by matching a specific channel pair to each task since under control conditions, it is the task which should cause a pair of channels to receive similar signals and thereby cause high correlation between phase and frequency. So theoretically, each task should have a channel pair with high phase/frequency linearity which exhibits less linearity for the other three tasks.

In this method, task phase arrays are constructed, the channel pair with highest correlation between phase and frequency is found for each task, phase/frequency relationships for the four channel pairs are plotted for every subject. Visual inspection of the plots gives categorizations of direction and maximum phase. Task dependent categorization differences are used to make rules.

8.1 Task Phase Arrays

Task phase arrays are obtained by the ultimate smoothing technique of averaging channel pair covariances across subjects in the same task group. It involves adding the array of smoothed channel pair covariances for each task's subject to a cumulative array of covariances. This results in one $43 \times 43 \times 252$ phase array for each task instead of 20 or so.

Averaging covariances across same-task subjects also produces remarkably smoothed phase/frequency relationships. For example, Figure 8.1 is the phase/frequency plot for channel pair (29,42) from the dot localization task phase array. The relationship is an almost perfectly negative linear one and it is this kind of ideal property that should allow distinctive channel pairs for each task to be found.

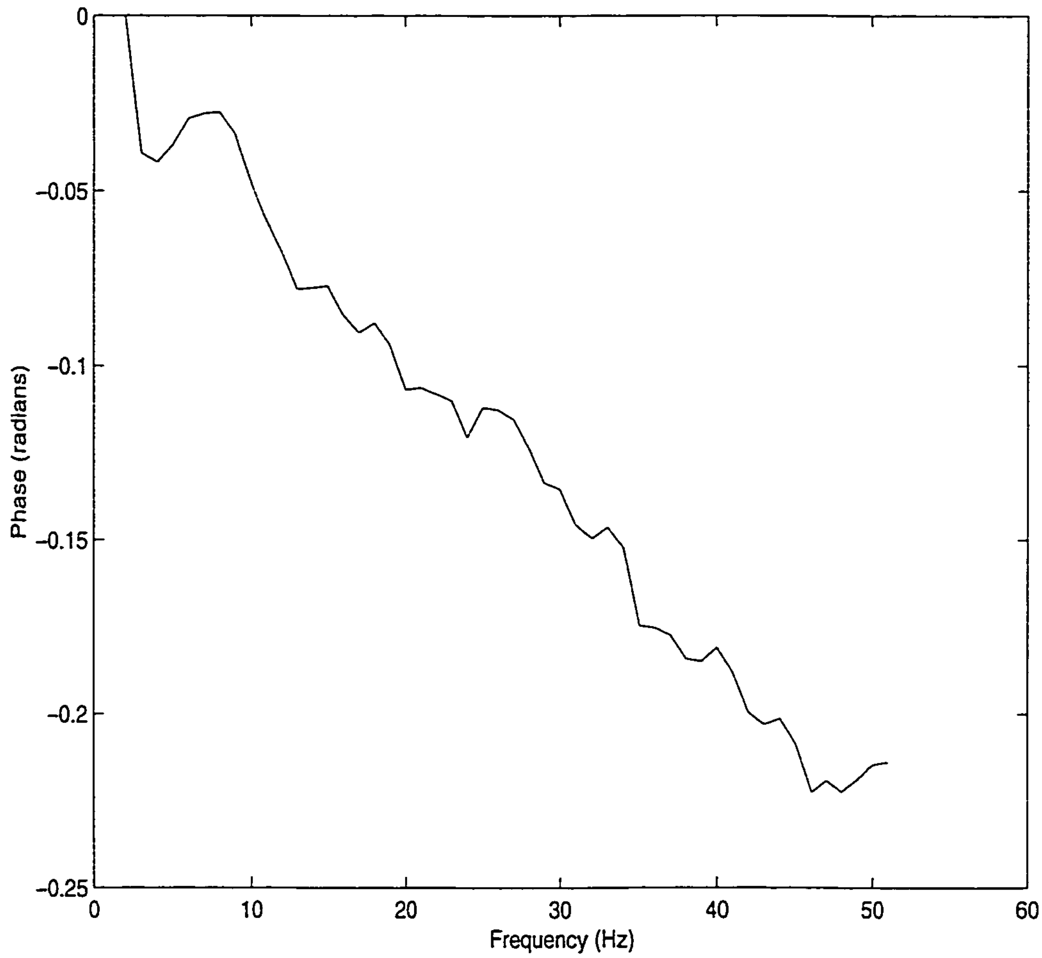

Figure 8.1: Phase/Frequency after Averaging Smoothed Covariances across Same-Task Subjects

This negative linear plot is for channel pair (29,42) from the dot localization task phase array and it shows that some tasks may produce strong phase/frequency linearity for some channel pairs.

8.2 Highest Correlation Task Channel Pairs

Next, the channel pair in each task's phase array with the highest correlation between phase and frequency must be found and for this, the same technique described in Section 7.3.1 is used. This reduces data from 43×43 channel pairs to only four (Table 8.1).

Specifically, taking CBP to be the covariance array after averaging covariances across same-task subjects, the phases are computed so that $PBP = \text{phase}(CBP)$. Then channels a, b ($a \neq b$) are determined so that if $Y = PBP_{a,b,\omega=2:50}$ and $X = \omega$ where $2 \leq \omega \leq 50$, then $|\text{correl}(Y, X)|$ would be a maximum. The procedure is repeated once for each task giving four such pairs (a, b).

Table 8.1: Highest Correlation Task Channel Pairs

Task	Channel Pair
Dot Localization	5,34
Word Finding	29,42
Eyes Open	6,34
Eyes Closed	9,34

8.3 Phase/Frequency Plots

Finally, the phase/frequency relationships of the four channel pairs are plotted for each subject, equally and independently of task. To express plotting procedures mathematically, we let CPD be the covariance matrix, after smoothing, for a typical subject. We denote eight channel numbers by a, b, c, d, e, f, g and h. We set the series $X = \omega = 2:50$, $Y_1 = \text{Phase}(CPD_{a,b,\omega=2:50})$, $Y_2 = \text{Phase}(CPD_{c,d,\omega=2:50})$,

$Y_3 = \text{Phase}(CPD_{e,f,\omega=2:50})$ and $Y_4 = \text{Phase}(CPD_{g,h,\omega=2:50})$. Then the relationships Y_1 vs. X, Y_2 vs. X, Y_3 vs. X and Y_4 vs. X are all plotted independently of the task for CPD.

8.4 Plot Characteristics

There are now 352 (88×4) channel pair phase/frequency plots to inspect and describe.

8.4.1 Plot Direction

A plot indicates whether the subject's task produces a link between the two channels. If the plot is just noise, it means that the signals received by the channel electrodes are totally dissimilar for that subject doing her particular task. A plot showing a linear relationship says that the two channels receive similar signals and are separated mainly by a lag when that subject performs her task. The direction of the linear relationship, whether positive or negative, tells which channel's signal is received first.

Therefore, every plot can be put into one of four directional categories.

(i) Noise is the dominant feature in the relationship, with the slope being neither significantly positive nor significantly negative. Figure 8.2 is an example where noise dominates and there is no apparent linear relationship between phase and frequency.

(ii) Noise is almost the dominant feature, but a positive or negative trend can be seen. (Figure 8.3)

(iii) A small amount of noise is present, but a positive slope dominates the relationship. Figure 8.4 shows a phase/frequency relationship that is clearly positive.

(iv) A small amount of noise is present, but a negative slope characterizes the relationship. Figure 8.5 is a case where the phase/frequency relationship is clearly negative.

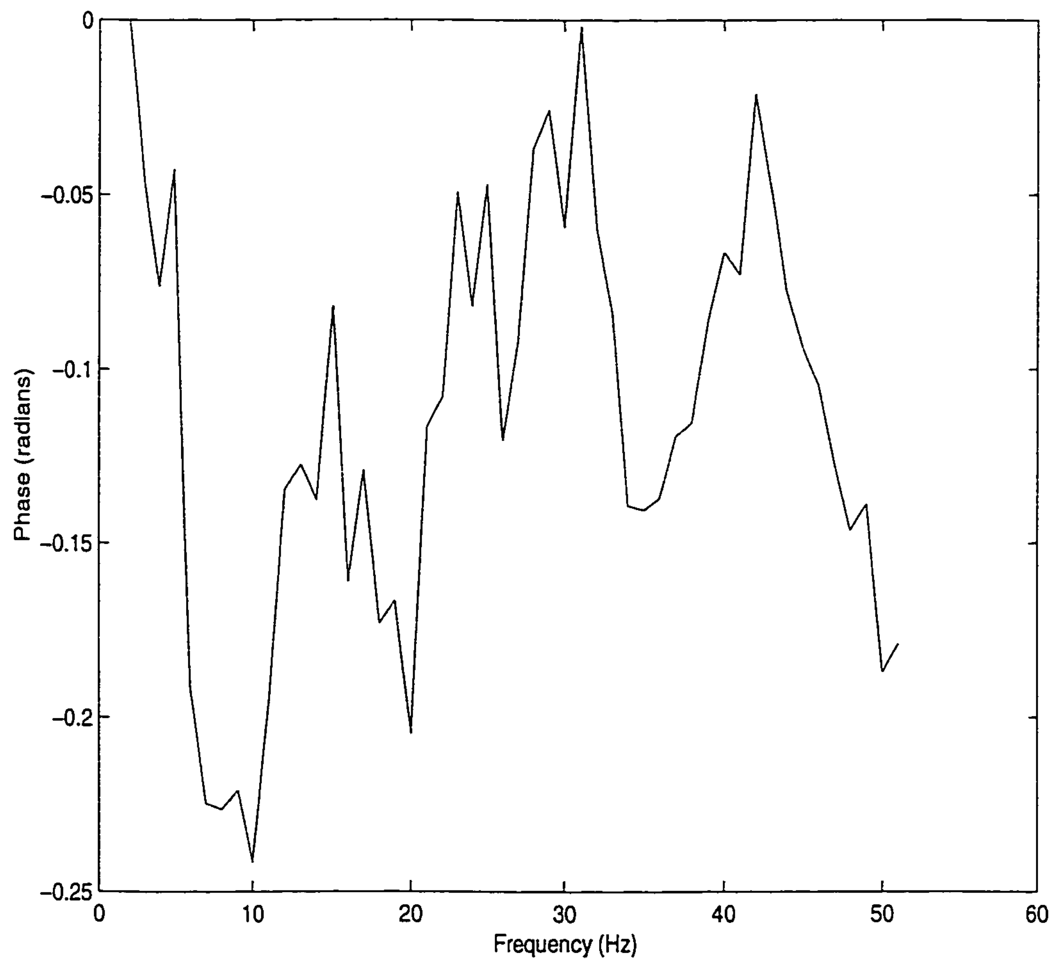

Figure 8.2: A Phase/Frequency Plot with Dominant Noise
(Channel pair (6,34) plotted for an eyes closed subject)

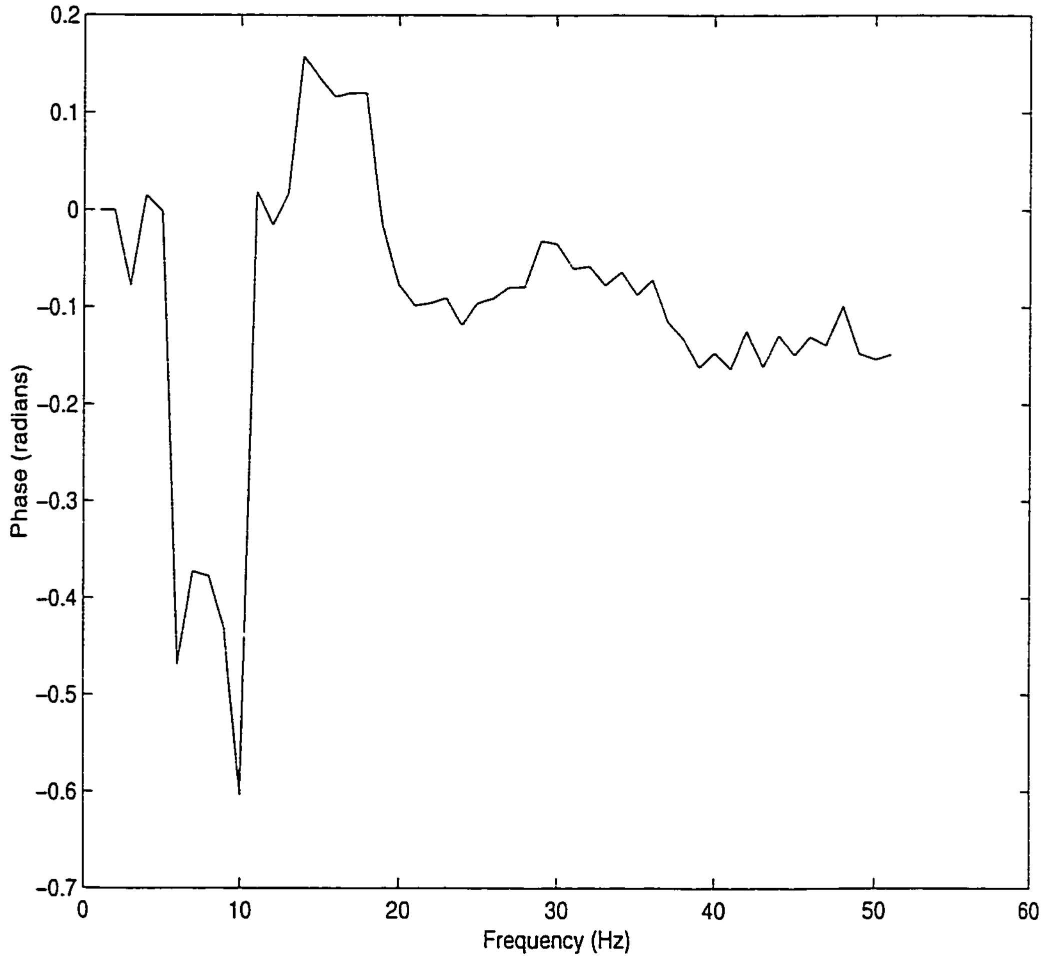

Figure 8.3: A Noisy Plot with a Negative Trend
Noise is the almost dominant feature in this plot for channel pair (29,42) for an eyes closed subject. However, if not for the large negative phase at the beginning, the plot would have shown a negative trend.

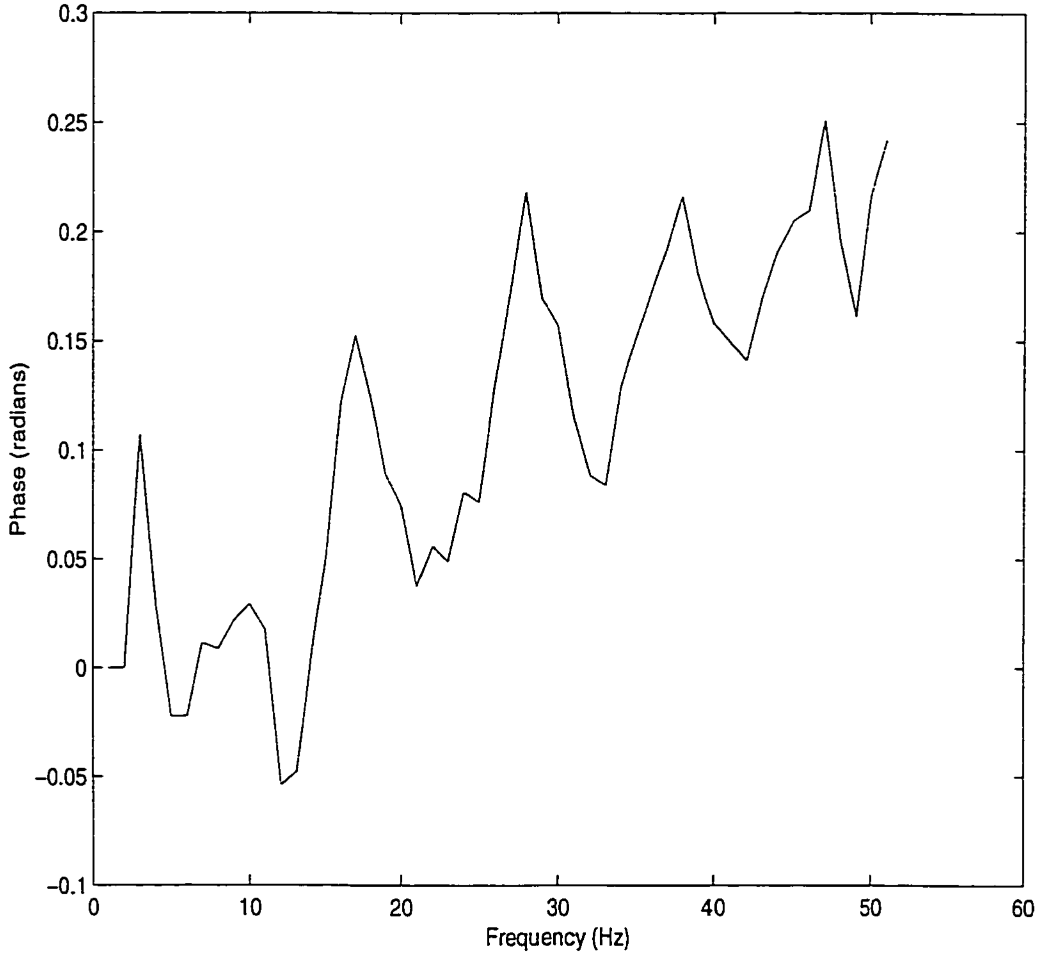

Figure 8.4: A Phase/Frequency Plot with a Clear Positive Relationship
(Channel pair (29,42) plotted for a dot localization subject)

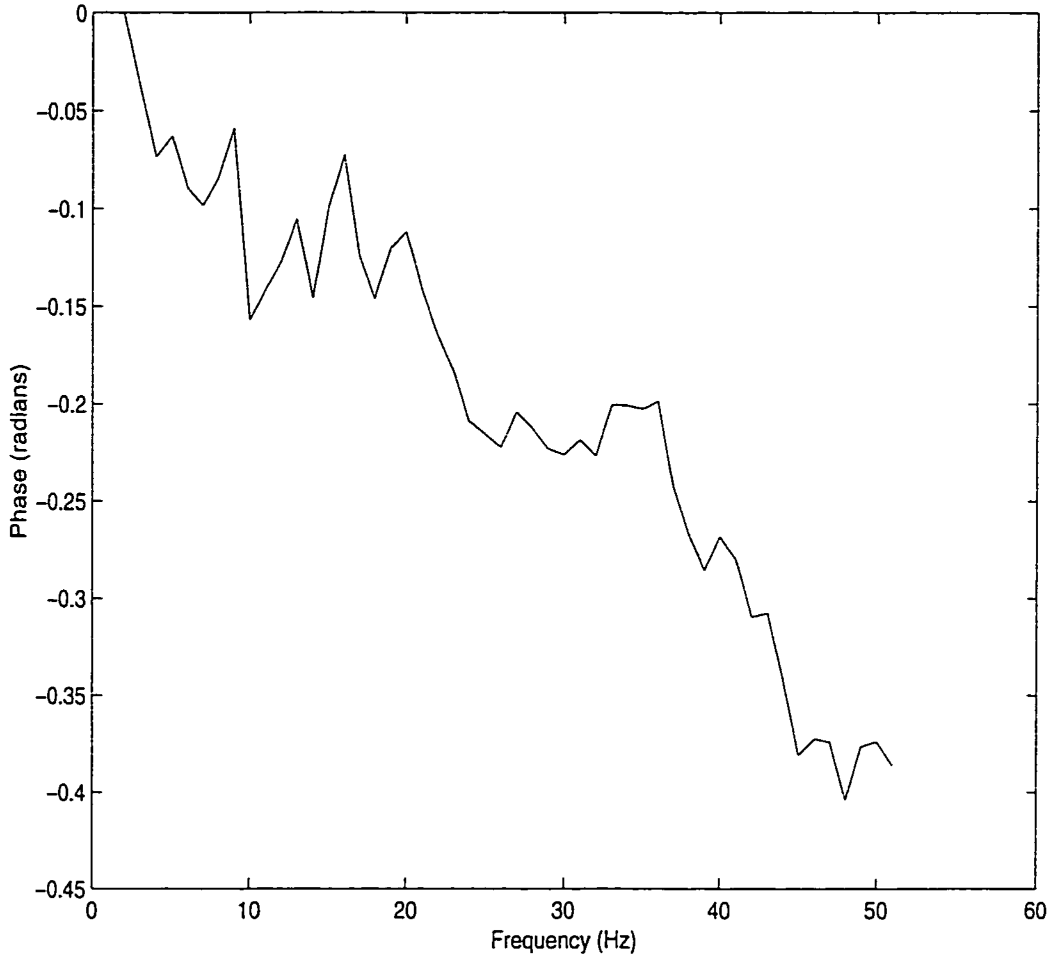

Figure 8.5: A Phase/Frequency Plot with a Clear Negative Relationship (Channel pair (29,42) plotted for a word finding subject)

8.4.2 Maximum Absolute Value of Phase

The size of signal lag affects how rapidly phase changes with respect to frequency and hence the maximum absolute value of the phase. Therefore, it is of special interest to see if there are any distinctive maximum phase values, exceptionally small or exceptionally large ones, in the four channel pair plots for each subject performing her task.

It was judged that maximum phase values between -0.2 radians and 0.2 radians were rare and hence distinctively small. Maximum phase values large enough to cause wrap arounds, that is, those above π or below $-\pi$ were also rare and distinctively large. Therefore, there are three maximum phase value categories: small, normal and large.

8.5 Ruling Out Conditions

With each of the four channel pair plots for every subject described in terms of direction and maximum phase value, an attempt can be made to find differences in the categorizations to classify subjects in the rule data set (Table 8.2).

Table 8.2: Phase/Frequency Descriptions, Highest Correlation Task Channel Pairs, Rule Data Set

Dot Localization Subjects

Channels (29, 42)	Channels (5,34)	Channels (6,34)	Channels (9,34)
negative	negative	negative	negative
negative	negative	negative	negative
negative	negative	negative, noise	negative
negative	negative	negative	negative
negative	negative	negative	negative
negative	negative	negative	negative
negative	negative	negative	negative
negative	negative	negative	negative
negative	negative	negative	negative
negative	negative	negative	negative
negative	negative	negative	negative

Word Finding Subjects

Channels (29,42)	Channels (5,34)	Channels (6,34)	Channels (9,34)
negative	negative	negative	negative
negative	negative	negative	negative
negative	negative	negative	negative
negative	negative	negative	negative
negative	negative	negative	negative
negative	negative	negative	negative
negative	negative	negative	negative
negative	negative	negative	negative
negative	negative	negative	negative
negative, small phase	negative	negative	negative

Eyes Open Subjects

Channels (29,42)	Channels (5,34)	Channels (6,34)	Channels (9,34)
negative	negative, small phase	negative	negative, small phase
negative	negative	negative	negative
negative	negative	negative	negative, small phase
negative	noise	negative	negative
negative	negative	negative	negative
negative	negative	negative	negative
negative	positive, small phase	negative	negative
negative	negative	negative	negative
negative	negative	negative	negative
negative	negative	negative, small phase	negative
negative	negative	negative	negative
negative	negative	negative	negative

Eyes Closed Subjects

Channels (29,42)	Channels (5,34)	Channels (6,34)	Channels (9,34)
negative	negative, small phase	negative	negative
negative	negative	negative	negative
negative	negative	negative	negative
negative	negative	negative	negative
negative	negative	negative	negative
negative	negative	negative	negative
negative	negative	negative	negative
negative	negative	negative	negative
negative	negative	negative	negative
negative	negative	negative	negative
negative	negative	negative	negative
negative	negative	negative	negative

It is obvious that phase/frequency behaves almost the same for all four tasks for all four channel pairs. Nevertheless, ruling out conditions could be developed for three tasks. Their reliabilities and efficiencies are on Table 8.3.

- Rule out dot localization if noise is dominant or if there is a small maximum phase for any of the four channel pairs.

- Rule out word finding if noise is dominant or almost dominant or if there is a small maximum phase for (5,34), (6,34) or (9,34).

- Rule out eyes closed if noise is dominant or almost dominant for (5,34), (6,34) or (9,34) or if there is a small maximum phase for either (6,34) or (9,34).

- No rule can be found for eyes open.

Table 8.3: Reliabilities and Efficiencies, Rule Data Set

Task	Reliability	Efficiency
Dot Localization	1	7/34
Word Finding	1	7/35
Eyes Open	_	_
Eyes Closed	1	5/33

With so little variation between tasks, reliabilities are 1 and efficiencies are low.

8.6 Test of Ruling out Conditions

Table 8.4 displays the plot categorizations for the test data set. The reliabilities and efficiencies of tests of the ruling out conditions are on Table 8.5.

Table 8.4: Phase/Frequency Descriptions, Highest Correlation Task Channel Pairs, Test Data Set

Dot Localization Subjects

Channels (29,42)	Channels (5,34)	Channels (6,34)	Channels (9,34)
negative	negative	negative	negative, noise
negative	negative	negative	negative
negative	negative	negative	negative
negative	negative	negative	negative
negative, small phase	negative	noise	negative
negative, small phase	negative	negative	negative
negative, small phase	negative	negative	negative
positive	negative	negative	negative
positive	negative	noise	negative
positive	negative	negative	negative
positive	negative	negative	negative

Word Finding Subjects

Channels (29,42)	Channels (5,34)	Channels (6,34)	Channels (9,34)
negative	negative	negative	negative
negative	negative	negative	negative
negative	negative	negative	negative
negative, noise	negative	negative	negative
negative, small phase	negative	negative	negative
negative, small phase	negative	negative	negative
negative	negative	negative	negative
positive	negative	negative	negative
positive, small phase	negative	negative	negative
positive, noise	negative	negative	negative

Eyes Open Subjects

Channels (29,42)	Channels (5,34)	Channels (6,34)	Channels (9,34)
noise	negative	negative	negative
negative	negative	negative	negative
negative	negative	negative	negative
negative	negative	negative	negative
negative, small phase	negative	negative, noise	negative
negative, small phase	negative	negative	negative
negative, small phase	negative	negative	negative
noise	negative	negative	negative
noise	negative	negative	negative
noise	negative	negative	negative
negative	negative	negative	negative

Eyes Closed Subjects

Channels (29,42)	Channels (5,34)	Channels (6,34)	Channels (9,34)
negative	negative	negative	negative
negative	negative	negative	negative
negative	negative	negative	negative
negative	negative	negative	negative
negative, small phase	negative	negative	negative
negative, noise	negative	negative	negative
negative, small phase	negative	negative	negative
noise	negative	negative	negative
noise	negative	negative	negative
noise	negative	negative	negative
noise	noise	noise	noise

Table 8.5: Reliabilities and Efficiencies, Test Data Set

Task	Reliability	Efficiency
Dot Localization	7/11	16/32
Word Finding	1	5/33
Eyes Open	-	-
Eyes Closed	10/11	4/32

Tests of the ruling out conditions for the three tasks are passed in that reliabilities and efficiencies sum to over 1. However, reliability falls for dot localization and the efficiencies for the other two tasks are low.

8.7 Highest Correlation Task Pairs without PCA Smoothing

It is possible that using all smoothing techniques on the covariances before averaging them and finding the highest correlation task channel pairs reduced task dependent phase/frequency differences too much. To see if smoothing was excessive, the entire procedure was repeated without PCA smoothing, since it did the least to reduce noise (Section 5.2.4).

Four different highest correlation task channel pairs were found (Table 8.6).

Table 8.6: Highest Correlation Task Channel Pairs without PCA

Task	Channel Pair
Dot Localization	9,15
Word Finding	2,21
Eyes Open	8,28
Eyes Closed	3,21

8.7.1 Ruling Out Conditions

Table 8.7 displays the plot categorizations for these new channel pairs for the rule data set.

Table 8.7: Phase/Frequency Descriptions, Highest Correlation Task Channel Pairs without PCA, Rule Data Set

Dot Localization Subjects

Channels (8,28)	Channels (2,21)	Channels (3,21)	Channels (9,15)
positive, noise	noise	noise	negative
positive	positive	positive	positive
noise	noise	negative	negative
positive	positive	noise	positive
positive	negative	noise	noise
noise	noise	positive	noise
noise	noise	noise	negative
positive	noise	noise	noise
positive	noise	noise	positive
positive	noise	positive	noise
noise	positive, noise	noise	noise

Word Finding Subjects

Channels (8,28)	Channels (2,21)	Channels (3,21)	Channels (9,15)
positive, noise	negative, noise	noise	negative
positive	noise	noise	positive
negative	negative	negative, noise	noise
positive, noise	positive, noise	noise	positive
noise	positive	positive, noise	negative
positive	positive, large phase	positive, noise	noise
positive	positive	positive, noise	positive
positive	positive, noise	positive	noise
positive	noise	noise	negative
noise	positive, noise	noise	noise

Eyes Open Subjects

Channels (8,28)	Channels (2,21)	Channels (3,21)	Channels (9,15)
negative, noise	noise	noise	positive
negative	positive	positive, noise	negative
negative, noise	noise	noise	noise
noise	noise	noise	negative
negative	negative, noise	noise	positive
negative	noise	noise	positive
noise	positive	positive	noise
noise	negative, noise	noise	noise
negative	noise	positive	negative
noise	noise	noise	noise
negative	noise	noise	positive
noise	noise	noise	negative

Eyes Closed Subjects

Channels (8,28)	Channels (2,21)	Channels (3,21)	Channels (9,15)
positive	positive, large phase	noise	negative
noise	negative, large phase	noise	noise
negative, noise	noise	noise	positive
noise	positive, large phase	noise	negative
negative, noise	noise, large phase	negative, large phase	positive
negative, noise	noise	negative	noise
negative, noise	positive, large phase	noise	positive
noise	negative	negative	noise
noise	positive, large phase	negative, large phase	negative
noise	noise	negative	positive
negative	noise, large phase	noise	positive
negative	negative	noise	noise

The dot localization task channel pair, (9,15), behaves randomly and has no task dependent pattern. Ruling out conditions can be developed for all four tasks from the other three channel pairs. Their reliabilities and efficiencies are on Table 8.8.

- Rule out dot localization if there is a large maximum phase for (2,21) or (3,21).
- Rule out word finding if there is a large maximum phase for (2,21) or (3,21).
- Rule out eyes open if there is a large maximum phase for (2,21) or (3,21) or if the direction for (8,28) is positive.
- Rule out eyes closed if the direction for (8,28) or (3,21) is positive or if the direction for (2,21) is positive at the same time that the maximum phase is not large.

Table 8.8: Reliabilities and Efficiencies, Rule Data Set

Task	Reliability	Efficiency
Dot Localization	1	8/34
Word Finding	11/12	7/35
Eyes Open	1	21/33
Eyes Closed	11/12	21/33

The phase/frequency plot categorizations for these four channel pairs produce ruling out conditions with the most promising reliabilities and efficiencies so far. Only two of the four efficiencies do not reach 0.5.

8.7.2 Test of Ruling Out Conditions

The ruling out conditions were applied to the test data set (Table 8.9) and the test reliabilities and efficiencies calculated (Table 8.10).

Table 8.9: Phase/Frequency Descriptions, Highest Correlation Task Channel Pairs without PCA, Test Data Set

Dot Localization Subjects

Channels (8,28)	Channels (2,21)	Channels (3,21)	Channels (9,15)
noise	noise	negative, noise	negative
noise	noise	noise	noise
positive	noise	positive	noise
negative	noise	noise	noise
negative	positive	negative, noise	negative
negative, noise	noise	noise	positive
negative	negative	noise	noise
negative	noise	noise	positive
negative	noise	noise	negative
negative	negative, noise	noise	noise
negative	noise	negative	positive

Word Finding Subjects

Channels (8,28)	Channels (2,21)	Channels (3,21)	Channels (9,15)
noise	noise	noise	negative
noise	noise	positive	noise
negative	positive, noise	positive, noise	positive
negative	negative, large phase	noise	noise
negative	noise	negative, noise	negative
negative	noise	noise	positive
negative, noise	noise	negative	noise
negative	noise	noise	positive
negative	negative, noise	negative, noise	negative
negative	noise, large phase	negative	noise

Eyes Open Subjects

Channels (8,28)	Channels (2,21)	Channels (3,21)	Channels (9,15)
noise	noise	noise	positive
negative, noise	noise	noise	noise
noise	noise	noise	negative
negative	negative	noise	noise
negative	noise, large phase	negative	negative
negative	negative	noise	positive
negative	negative	negative	positive
noise	negative, large phase	noise	noise
positive, noise	positive, noise	negative	negative
negative	noise	noise	noise
negative	noise	noise	negative

Eyes Closed Subjects

Channels (8,28)	Channels (2,21)	Channels (3,21)	Channels (9,15)
negative	noise, large phase	positive, large phase	noise
noise	negative	noise	positive
noise	positive, large phase	negative, noise	negative
negative, noise	negative	negative	noise
negative	noise	negative	negative
noise	positive, large phase	noise	positive
negative	noise	noise	positive
negative	positive, large phase	noise	noise
negative	noise	noise	negative
negative	negative	negative, noise	noise
negative	negative, noise	negative, noise	negative

Table 8.10: Reliabilities and Efficiencies, Test Data Set

Task	Reliability	Efficiency
Dot Localization	1	8/32
Word Finding	8/10	6/33
Eyes Open	8/11	7/32
Eyes Closed	10/11	4/32

The test is failed for word finding and eyes open and while it is passed for the other two tasks, their efficiencies are far below 0.5.

8.8 Discussion

No direct task identification rules can be found.

Using categorizations of highest correlation task channel pair phase/frequency plots produces ruling out conditions for three tasks which pass tests when PCA is included in covariance smoothing. However, reliabilities fluctuate in tests and efficiencies are never above 0.5.

Without PCA smoothing, the highest correlation task channel pairs produce ruling out conditions for four tasks, only two of which pass tests. Efficiencies and reliabilities both fall when rules are tested.

When PCA smoothing is included, phase/frequency does not vary enough between tasks. When it is omitted, phase/frequency varies too much within tasks. This indicates either a low between-task to within-task variation ratio or a difficulty in finding the right level of smoothing.

Chapter 9

Numerical Classification Using Highest Correlation Task Channel Pairs Correlation/Slope Vectors

With numerical classification, the complexity of using several phase/frequency properties to discriminate can be reduced by taking the "optimal" linear combinations of correlation and least squares slope. A numerical procedure using correlation/slope vectors of phase versus frequency for distinctive channel pairs was judged a reasonable route to follow since correlation is high when the signals of two channels are close to identical and least squares slope is high when their lag is large.

The four highest correlation task channel pairs computed by across subject averaging after all smoothing were chosen. These are (29,42), (5,34), (6,34) and (9,34) (Section 8.2).[1] Correlation/slope vectors are computed for these four channel pairs for each subject's phase array. Linear combinations which maximize the statistical distance between correlation/slope vectors for each pair of tasks are calculated. Discrimination scores using these linear combinations are obtained for each subject's

[1]The four highest correlation task channel pairs found after omitting PCA (Section 8.7) were also investigated. The procedure was abandoned when it was found that the statistical distances from the maximizing linear combinations were less than 1.

phase array and boundaries found in these are used for classification.

9.1 Correlation/Slope Vectors

Taking channel pairs (29,42), (5,34), (6,34) and (9,34) in order, the correlation coefficients between phase and frequency for the 45 subjects in the rule data set are calculated. Using the same four channel pairs in the same order, the least squares slopes of phase/frequency for each subject are determined. The result is a vector of length 8 for each subject. Correlation coefficients and least squares slopes are computed by Matlab.

9.2 Linear Combinations

To use correlation/slope vectors in discrimination, they must be converted to discrimination scores via linear combinations which maximize the statistical distance (measured by the number of standard deviations) between pairs of tasks. These linear combinations pool correlation and slope in the optimal way for task separation.

The first step in calculating the linear combinations for each pair of tasks is to compute the mean vector and covariance matrix for each task. Suppose A is the task and $x_i (i = 1..n)$ are the correlation slope vectors for each subject in that task group. Then the mean vector for task A is $\overline{x} = \sum_{i=1}^{n} x_i/n$ and the covariance matrix is $\Sigma = (\sum_{i=1}^{n} x_i x_i\prime - n\overline{x}\overline{x}\prime)/(n-1)$. Taking each pair of tasks, we use their mean vectors and covariance matrices in determining the maximizing linear combinations. For tasks A and B, let us denote the mean correlation/slope vector for task A as $\overline{x}_1$ and the one for task B as $\overline{x}_2$. If Σ_1 and Σ_2 are the covariance matrices for the two tasks respectively, then the linear combination a is computed so that $a'(\overline{x}_1 - \overline{x}_2)/\sqrt{a'\Sigma_p a}$ is a maximum. Here, Σ_p is the pooled covariance matrix, that is, $\Sigma_p = ((n_1 - 1)\Sigma_1 + (n_2 - 1)\Sigma_2)/(n_1 + n_2 - 2)$.

To compute the linear combinations with Matlab, the mean vectors and covariance matrices were fed into a standardizing function, $a = \sqrt{\Sigma_p^{-1}}(\overline{x}_1 - \overline{x}_2)$ (Table 9.1).

Table 9.1: Linear Combinations for the Six Pairs of Tasks, Rule Data Set

Task Pair	Statistical Distance	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8
DL vs. WF	$\sqrt{1.7412}$	0.7724	-0.6808	-0.1718	0.0036	0.1973	-0.6871	0.2099	-0.3106
DL vs. EO	$\sqrt{2.2446}$	0.5664	-0.9061	0.039	0.3804	0.6392	-0.1865	0.4333	0.5704
DL vs. EC	$\sqrt{2.2050}$	0.7459	-0.2899	-0.8458	0.4001	0.2154	-0.1257	0.7223	0.3243
WF vs. EO	$\sqrt{1.6603}$	-0.2853	0.26	0.32	0.3193	0.9422	0.4166	-0.077	0.4938
WF vs. EC	$\sqrt{2.3757}$	0.0275	0.2523	-0.6986	0.5114	-0.1767	0.6306	-0.4063	0.9837
EO vs. EC	$\sqrt{1.3367}$	0.3464	0.1349	-0.7289	0.2638	-0.5302	-0.3407	0.3181	0.315

Looking at the statistical distances, a potential problem is evident. All tasks are within two standard deviations of each other and this may be too low for successful discrimination. Nevertheless, since the distances are greater than 1, the procedure was pursued further.

9.3 Discrimination Scores

There are six pairs of tasks and hence six linear combinations. Therefore, six discrimination scores for each subject can be obtained by multiplying each linear combination by her correlation/slope vector. If a subject's correlation/slope vector is x and the linear combinations are $a_1 \ldots a_8$, then her six discrimination scores are $a_1 x \ldots a_8 x$. (Table 9.2).

Table 9.2: Discrimination Scores, Rule Data Set

Dot Localization Subjects

DL vs. WF	DL vs. EO	DL vs. EC	WF vs. EO	WF vs. EC	EO vs. EC
0.04634015	-0.08250632	-0.12532771	-0.05514418	-0.1752458	-0.10426312
0.03740963	-0.11478162	-0.05950744	-0.08838079	-0.09681692	-0.03115176
-0.1994366	-0.25971442	-0.32057932	0.01669283	-0.14491872	-0.17533578
-0.0780749	-0.1572536	-0.22697932	-0.0057663	-0.1582785	-0.14384656
-0.08504674	-0.20743313	-0.21311482	-0.05259319	-0.14957234	-0.11929882
-0.07640747	-0.16662647	-0.2962152	-0.0071577	-0.24146763	-0.20740725
-0.06456947	-0.1864551	-0.14214382	-0.05879117	-0.09312362	-0.06750969
0.0652262	-0.07976861	-0.0399451	-0.08224473	-0.10418462	-0.03246093
-0.13621821	-0.20597494	-0.25139258	0.00138719	-0.13245839	-0.14197674
0.00615142	-0.08633677	-0.03318411	-0.03858385	-0.03292099	-0.01096166
0.09199858	-0.01520269	0.0145793	-0.04827105	-0.06171948	-0.00765834

Word Finding Subjects

DL vs. WF	DL vs. EO	DL vs. EC	WF vs. EO	WF vs. EC	EO vs. EC
0.0328542	-0.10434882	-0.1091806	-0.06627627	-0.1475322	-0.08100912
0.01231476	-0.0560102	-0.136881	0.00133191	-0.14510883	-0.11309405
0.01329217	-0.10598371	-0.12065754	-0.05630209	-0.1415292	-0.08592339
0.07677373	-0.07090328	-0.00360212	-0.08892104	-0.07713436	-0.00465094
0.17860274	0.00064333	0.07320394	-0.12481755	-0.09727397	0.02090689
-0.00844296	-0.06540366	-0.1108482	0.00679603	-0.09364872	-0.08082268
0.0226823	-0.14135701	-0.07704782	-0.09723093	-0.11167069	-0.01876472
0.07755163	-0.0665836	-0.11387811	-0.07005974	-0.19300519	-0.10266511
0.09225789	-0.0341798	-0.01885086	-0.06246032	-0.1037429	-0.03338897
0.45953665	0.24011398	0.35448873	-0.19802462	-0.07318045	0.14523335

Eyes Open Subjects

DL vs. WF	DL vs. EO	DL vs. EC	WF vs. EO	WF vs. EC	EO vs. EC
-0.13924976	-0.13796323	-0.24872244	0.07048268	-0.11261041	-0.15739443
0.11938758	-0.04677854	-0.00821642	-0.10358867	-0.12651828	-0.02624134
-0.04935546	-0.08169725	-0.07740932	0.01759003	-0.02046683	-0.03848884
0.03969069	-0.06774532	-0.10031217	-0.03980771	-0.14030008	-0.08493924
0.09444409	-0.07590081	-0.00378445	-0.11073457	-0.10003387	-0.0095822
0.0959252	0.00541344	0.06925189	-0.04215706	-0.00554722	0.03555153
0.02057579	-0.105112	-0.06715825	-0.06224624	-0.0875592	-0.03993448
0.22722847	0.14835651	-0.17738423	0.00737687	-0.39357416	-0.25011008
0.2225121	0.03458064	0.11163331	-0.13738076	-0.09806805	0.03691981
-0.03734508	-0.11145853	-0.27100994	0.0119048	-0.24563347	-0.20842916
0.02201572	-0.06826972	-0.07866811	-0.02847257	-0.09624146	-0.06106578
0.07907932	-0.03644	-0.00184335	-0.060049	-0.07238267	-0.01321994

Eyes Closed Subjects

DL vs. WF	DL vs. EO	DL vs. EC	WF vs. EO	WF vs. EC	EO vs. EC
0.0630292	-0.08241288	0.05304416	-0.10683451	-0.01053346	0.05130568
0.15909397	-0.01227949	0.07485762	-0.12113002	-0.07625676	0.03081832
0.12096745	0.02798043	0.07083797	-0.05520711	-0.03133993	0.02809821
0.05427011	-0.08710768	-0.04722524	-0.07888892	-0.10105178	-0.03425026
0.2162104	0.02968308	0.07812152	-0.13576367	-0.12844582	0.0111587
0.14877004	0.0299816	0.05759079	-0.06774077	-0.07464097	0.00608491
-0.25814588	-0.25271005	-0.22523613	0.06023942	0.02253308	-0.07785176
0.28487076	0.0945724	0.16223588	-0.1479554	-0.10536056	0.05231225
0.12455934	0.14061457	0.19310782	0.04275602	0.12290406	0.10719466
-0.00774265	-0.10798226	-0.14084411	-0.03117283	-0.13817502	-0.09951778
0.13701077	0.03781904	0.08606019	-0.0538537	-0.02890324	0.03284909
0.04662686	-0.10218239	-0.07193049	-0.08579107	-0.12769397	-0.05372947

An attempt is now made to find boundaries that can separate some tasks from others, both by direct identification and by ruling out conditions.

9.4 Direct Task Identification Rules

9.4.1 Rules

The rules that can be found to identify tasks directly, together with their reliabilities and efficiencies, are on Table 9.3.

Table 9.3: Direct Rules, Reliabilities and Efficiencies, Rule Data Set

Task	Identify Task If:	Reliability	Efficiency
Dot Localization	DL vs. WF < 0	28/34	6/11
Word Finding	–	–	–
Eyes Open	WF vs. EO > 0	27/33	4/12
Eyes Closed	DL vs. EC > 0	28/33	8/12

There are direct identification rules for dot localization, eyes open and eyes closed, with zero being a convenient separation boundary. Reliabilities are all above 0.8 and efficiencies pass the 0.5 mark for dot localization and eyes closed. However, word finding can not be identified.

9.4.2 Testing

The test data set is processed to obtain correlation/slope vectors and, using the same linear combinations, discrimination scores are calculated (Table 9.4). The rules are then tested on these discrimination scores and reliabilities and efficiencies found. (Table 9.5).

Table 9.4: Discrimination Scores, Test Data Set

Dot Localization Subjects

DL vs. WF	DL vs. EO	DL vs. EC	WF vs. EO	WF vs. EC	EO vs. EC
0.13949662	0.1061359	0.25044646	-0.02018241	0.15429931	0.16262546
0.01538063	-0.11279548	-0.0874859	-0.06439373	-0.10761787	-0.05559959
0.07767544	-0.06627686	-0.03872939	-0.08536387	-0.11793446	-0.03861918
0.1378226	-0.03594302	0.00721279	-0.1143922	-0.13077816	-0.02032874
0.19191539	0.12998442	-0.28074275	0.03000045	-0.4742912	-0.33367076
0.06732645	-0.10146736	-0.02981117	-0.10887445	-0.10350337	-0.0195963
0.06330665	-0.10104954	-0.02901742	-0.10590859	-0.09906528	-0.02058638
1.28550592	0.92663366	1.30364175	-0.45114409	0.13270108	0.63034442
1.0048371	0.67539459	0.72631517	-0.35368628	-0.21932937	0.24120445
1.54050558	1.00199746	1.37610563	-0.61882233	-0.06521475	0.61415897
1.2600336	0.88214471	1.12530202	-0.44200206	-0.03743356	0.48739536

Word Finding Subjects

DL vs. WF	DL vs. EO	DL vs. EC	WF vs. EO	WF vs. EC	EO vs. EC
0.09371343	-0.05432871	-0.00327785	-0.08917681	-0.09190135	-0.01241565
0.29555165	0.08823464	0.18373438	-0.16797142	-0.09634323	0.07041719
0.16730002	-0.01266688	0.04268368	-0.12474483	-0.1216839	-0.00128437
0.4322862	0.20244058	0.22464448	-0.18713897	-0.18728343	0.04614669
0.17860081	-0.00653081	0.0044061	-0.14217171	-0.1860209	-0.03946136
0.40781372	0.1569476	0.29074679	-0.21815929	-0.09721159	0.12129415
1.30005064	0.86595156	1.11545326	-0.49596032	-0.10029715	0.47217531
1.32690114	0.86772753	1.19309037	-0.52995458	-0.0463226	0.5357386
1.15052455	0.70290951	1.01703332	-0.49332171	-0.06222798	0.46130653
1.04284573	0.76081755	0.95720417	-0.33586439	0.00642177	0.41928643

Eyes Open Subjects

DL vs. WF	DL vs. EO	DL vs. EC	WF vs. EO	WF vs. EC	EO vs. EC
0.54424023	0.32809749	0.42102204	-0.20188399	-0.07920188	0.16381601
0.11978156	-0.01749697	0.04476856	-0.0882386	-0.06710598	0.01349466
0.18329214	0.00380319	0.10976976	-0.13363112	-0.0643676	0.05109998
0.12577051	-0.01992347	-0.03228088	-0.08014299	-0.15356713	-0.0561056
0.12297309	0.03713617	-0.31533831	0.01329905	-0.44906564	-0.3255386
0.05584647	-0.08894515	-0.05437237	-0.08716882	-0.11603122	-0.04111126
0.20606195	0.00795523	0.10843762	-0.14761005	-0.09028949	0.04323398
0.97479828	0.5991806	0.85747284	-0.40537393	-0.05014161	0.38652483
0.58664654	0.32849215	0.35606408	-0.23121538	-0.19926312	0.09457386
0.88552489	0.52272227	0.75157772	-0.37900457	-0.07731547	0.33042089
0.9027088	0.63661496	0.71349176	-0.28932392	-0.11967708	0.26434749

Eyes Closed Subjects

DL vs. WF	DL vs. EO	DL vs. EC	WF vs. EO	WF vs. EC	EO vs. EC
0.6618809	0.37915509	0.56593948	-0.29021	-0.05242961	0.25305346
0.00239488	-0.01583642	-0.05801745	0.03185761	-0.04440115	-0.05034965
0.65665656	0.3701652	0.49871542	-0.27867864	-0.11846169	0.19595602
0.17330641	0.01686558	0.05281149	-0.10331554	-0.11141335	-0.00087367
0.56824711	0.3072239	0.38682702	-0.23749761	-0.14874166	0.13182475
0.70549761	0.37635147	0.59890742	-0.3354286	-0.06646944	0.27299259
0.11976783	-0.03970896	0.02620648	-0.10418327	-0.09072605	0.0017902
0.97918035	0.67032768	0.83815952	-0.34011863	-0.06643909	0.34817091
1.03626996	0.65736591	0.88740443	-0.40901266	-0.07927023	0.38156022
0.64734915	0.35351327	0.50675328	-0.28510281	-0.10164821	0.20939852
-0.05652053	0.02800472	0.03358657	0.08392047	0.11249911	0.03654213

Table 9.5: Reliabilities and Efficiencies, Test Data Set

Task	Reliability	Efficiency
Dot Localization	31/32	0
Word Finding	–	–
Eyes Open	29/32	1/11
Eyes Closed	9/32	10/11

Generally, the test is failed because, in the cases of dot localization and eyes open, the probability of a Type I error is greater than the probability of a correct identification. The eyes closed rule barely passes.

9.5 Ruling Out Conditions

9.5.1 Rules

Ruling out conditions developed from boundaries in the rule data set discrimination scores are more complicated than direct identification. It is helpful to follow the process through three tables (9.6, 9.7 and 9.8).

Table 9.6: Initial Ruling Out Conditions, Reliabilities and Efficiencies, Rule Data Set

Task	Rule Out Task If:	Reliability	Efficiency
Dot Localization	DL vs. EO > -0.078	10/11	22/34
Word Finding	DL vs. WF < 0	9/10	11/35
Eyes Open	DL vs. EO < -0.11	10/12	9/33
Eyes Closed	WF vs. EC < -0.13	11/12	13/33

The reliability standard of 0.8 is met, but efficiencies are well below 0.5. There is room to weaken conditions and therefore raise efficiencies for word finding and eyes closed.

Table 9.7: Additional Ruling Out Conditions, Reliabilities and Efficiencies, Rule Data Set

Task	Rule Out Task If:	Reliability	Efficiency
Word Finding	WF vs. EC > -0.07	1	9/35
Eyes Closed	DL vs. EC < -0.2	11/12	7/33

When weaker conditions are created by combining the two sets of ruling out conditions, reliabilities are still above 0.8 and efficiencies for word finding are raised to over 0.5. This means that word finding and dot localization can be ruled out and that the passive tasks can be separated from the actives. There is no way to raise the efficiency of eyes open and eyes closed.

Table 9.8: Combined Ruling Out Conditions, Reliabilities and Efficiencies, Rule Data Set

Task	Rule out Task If:	Reliability	Efficiency
Dot Localization	DL vs. EO > -0.078	10/11	22/34
Word Finding	DL vs. WF < 0 or WF vs. EC > -0.07	9/10	19/35
Eyes Open	DL vs. EO < -0.11	10/12	9/33
Eyes Closed	WF vs. EC < -0.13 or DL vs. EC < -0.2	10/12	13/33

9.5.2 Testing

Here are the results of tests of the ruling out conditions.

Table 9.9: Reliabilities and Efficiencies, Test Data Set

Task	Reliability	Efficiency
Dot Localization	3/11	31/32
Word Finding	7/10	12/33
Eyes Open	1	1/32
Eyes Closed	10/11	8/32

In general, the test results of weaker ruling out conditions are far from satisfactory, especially with the low efficiencies for word finding, eyes open and eyes closed and with the sharp fall in reliability for dot localization. However, the test is passed because reliabilities and efficiencies sum to over 1 in all cases.

9.6 Discussion

As predicted, the lack of statistical distance between tasks causes problems in developing both direct task identification rules and ruling out conditions.

The failed test of direct rules again points to the fact that phase/frequency does not differ enough between tasks for direct four-way classification.

In addition, simple ruling out conditions are not to be had because they are too strong to have good efficiencies. However, this method does produce a four-way classification which passes tests, albeit poorly.

Chapter 10

Numerical Classification Using Highest Slope Task Channel Pairs Correlation/Slope Vectors

In one more trial to find numerical discrimination, a different set of task channel pairs was investigated. Instead of highest correlation, the highest least squares phase/frequency slope was used to find a channel pair for each task. This method is suggested by the fact that slope is high when lag between the signals of two channels is large.

10.1 Highest Least Squares Slope Task Channel Pairs

The task phase arrays obtained from the smoothed channel pair covariances averaged across same-task subjects (Section 8.1) are transformed to least squares slopes arrays by using Matlab and the channel pair corresponding to the highest least squares slope in absolute value is obtained.

The channel pairs found are (12,36) for dot localization, (32,43) for word finding, (31,40) for eyes open and (32, 41) for eyes closed.

10.2 Linear Combinations and Discrimination Scores

Using these four channel pairs, correlation/slope vectors are computed for each of the 45 subjects in the rule data set and converted to discrimination scores via linear combinations (Tables 10.1 and 10.2).

Table 10.1: Linear Combinations for the Six Pairs of Tasks, Rule Data Set

Task Pair	Statistical Distance	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8
DL vs. WF	$\sqrt{1.8976}$	0.0075	-1.0602	0.2115	-0.3836	-0.3557	-0.0990	0.1733	-0.6444
DL vs. EO	$\sqrt{1.1794}$	0.7449	-0.2952	-0.1189	-0.2689	0.3323	-0.4414	0.2522	-0.2864
DL vs. EC	$\sqrt{1.6286}$	0.1453	-0.6515	-0.2811	-0.5682	-0.2683	-0.2047	0.7932	-0.1952
WF vs. EO	$\sqrt{1.3164}$	0.6620	0.7427	-0.2220	0.3203	0.2842	-0.2951	-0.0155	-0.0813
WF vs. EC	$\sqrt{1.4432}$	0.0675	0.4160	-0.5050	-0.0458	0.0695	-0.4280	0.5925	0.6851
EO vs. EC	$\sqrt{1.7098}$	-0.5804	-0.4592	-0.2260	-0.3729	-0.2161	0.7202	0.4914	0.4064

Here, statistical distances are even smaller than those for the highest correlation task pairs in Table 9.1, being all less than $\sqrt{2}$. Nevertheless, they are still all greater than 1 and so the method was pursued.

Table 10.2: Discrimination Scores, Rule Data Set

Dot Localization Subjects

DL vs. WF	DL vs. EO	DL vs. EC	WF vs. EO	WF vs. EC	EO vs. EC
-1.58964826	-0.88738666	-1.04086399	0.72413975	0.67689121	-0.07099764
-0.26065446	0.43643385	-0.32670491	0.6740329	-0.12462993	-0.79752426
-1.29709886	-0.33866422	-1.10318504	1.04260876	0.28305767	-0.80140968
1.20104423	0.18557043	0.77432834	-1.04503162	-0.51924195	0.61808511
0.24258088	-0.22299338	0.27273252	-0.4886493	0.05198178	0.54007958
0.11610264	-0.6079024	-0.32421852	-0.5437147	-0.3453629	0.21465838
-0.56241376	-0.70442632	-0.72717034	0.01012275	-0.08784447	-0.13321197
-0.78021199	-0.91054017	-0.93710783	0.05642071	-0.06216525	-0.16933091
-0.73374349	-0.30589671	-0.39461169	0.44241751	0.45797691	-0.06295462
-0.85490885	-0.62483477	-0.47098166	0.26146344	0.49143191	0.13648159
-0.46889407	-0.66883857	-0.75004487	-0.00259465	-0.19121062	-0.21525007

Word Finding Subjects

DL vs. WF	DL vs. EO	DL vs. EC	WF vs. EO	WF vs. EC	EO vs. EC
1.54534746	0.20318743	0.96266207	-1.28537764	-0.71844549	0.66813951
0.62994777	-0.37357789	0.53107615	-0.92482353	-0.23982585	0.792805
0.74979195	0.70888673	0.91529626	-0.22071799	0.06646811	0.31321466
1.49318336	0.22333263	0.98372192	-1.32327474	-0.62138404	0.81522058
0.59390775	-0.18625746	0.233897	-0.68742332	-0.26959165	0.40974571
-0.36690033	-0.57305723	-0.64542523	-0.01603432	-0.19819108	-0.20618493
0.79320873	-0.69536725	0.00496013	-1.16706369	-0.62173299	0.56713213
0.52937793	-0.45660136	-0.18474013	-0.74532088	-0.57013874	0.18117411
-0.81152379	-0.77842381	-0.9806984	0.23254751	-0.06601166	-0.3472079
1.05059919	1.38172141	1.30650455	-0.04210413	0.07773542	0.19224198

Eyes Open Subjects

DL vs. WF	DL vs. EO	DL vs. EC	WF vs. EO	WF vs. EC	EO vs. EC
0.58648545	0.10111122	0.68976771	-0.52969583	-0.07240391	0.56727035
-1.01992999	0.36082593	-0.53185174	1.28826118	0.49761146	-0.86379152
-0.37759799	0.41839362	-0.21517027	0.67040328	-0.07179179	-0.6543109
0.11392635	0.78123315	0.2176732	0.4912242	-0.00794768	-0.47420681
-0.1938501	-0.78318925	-0.52859081	-0.41643197	-0.26156707	0.14191452
0.06770754	-0.08116369	-0.2794136	-0.04249787	-0.31539865	-0.26215196
-0.97400978	-0.89404581	-0.69936959	0.19640825	0.43128226	0.12841582
1.21307293	0.50490881	1.2269812	-0.85971014	-0.06547869	0.79647356
0.63038883	0.5283779	0.08813618	-0.00851999	-0.49952992	-0.45138758
-1.57800821	0.07243851	-0.97533534	1.53357583	0.72250892	-0.94315751
0.24386515	-0.4052817	-0.22228287	-0.50731006	-0.43731808	0.09981628
-0.12999127	0.52056149	0.01072413	0.59464856	0.32330538	-0.41706936

Eyes Closed Subjects

DL vs. WF	DL vs. EO	DL vs. EC	WF vs. EO	WF vs. EC	EO vs. EC
0.53562832	0.49116308	0.66711754	-0.19086186	-0.0381344	0.22348819
-0.05470861	-0.61001532	-0.39819231	-0.39375986	-0.28532142	0.10333918
1.07110203	0.62834066	1.08424576	-0.58793471	-0.08313226	0.53616593
0.63666828	0.18299342	0.80157771	-0.58362586	0.1127016	0.71632984
-1.44457699	-0.42495647	-0.96546852	1.0607732	0.58811935	-0.55164156
-0.40098491	-0.31807908	-0.48316821	0.19724309	0.08652384	-0.17316958
1.34262166	0.27021607	0.86580598	-1.07366599	-0.55059748	0.58190569
0.79296354	-0.4391507	-0.10397397	-0.97484344	-0.66662126	0.26006818
-0.61075364	-0.02710197	0.07914751	0.39865602	0.50434258	0.14234355
0.35216856	-0.66860982	0.18969218	-0.9098936	-0.23673203	0.68581537
0.85386518	1.073446	0.78702151	0.04170058	-0.14155427	-0.14882972
-0.06389174	-0.510788	-0.07431643	-0.40890284	0.03846001	0.42541307

10.3 Direct Task Identification Rules

10.3.1 Rules

The discrimination scores are examined for boundaries to directly identify tasks.

Table 10.3: Direct Rules, Reliabilities and Efficiencies, Rule Data Set

Task	Identify Task If:	Reliability	Efficiency
Dot Localization	DL vs. EC < -0.7	31/34	5/11
Word Finding	WF vs. EC < -0.5	32/35	4/10
Eyes Open	EO vs. EC < -0.25	29/33	7/12
Eyes Closed	-	-	-

Three tasks can be identified with reliabilities above 0.8, their numerical boundaries being straightforward. Efficiencies are below 0.5 for dot localization and word finding. No direct rule can be found to identify eyes closed.

10.3.2 Testing

Table 10.4 contains the discrimination scores for the test data set. These are used to test the direct rules for reliability and efficiency (Table 10.5).

Table 10.4: Discrimination Scores, Test Data Set

Dot Localization Subjects

DL vs. WF	DL vs. EO	DL vs. EC	WF vs. EO	WF vs. EC	EO vs. EC
1.14196046	0.45914141	1.12903623	-0.81402259	-0.15164918	0.77496081
0.52875412	0.94908013	0.67557442	0.24132212	0.13172133	-0.10976693
-0.41609655	-0.50797006	-0.4209065	-0.00121448	0.06763516	0.03054971
0.27668776	0.12031959	-0.2246327	-0.02758801	-0.39888907	-0.34165093
0.84662423	-0.35037449	0.44589043	-1.10607521	-0.49005665	0.70762543
-0.14100578	-0.31635408	-0.00134876	-0.16831051	0.19356582	0.33164767
-0.28394152	-0.42700124	-0.0554954	-0.1795719	0.23818915	0.36670665
-0.32251145	-0.29526625	-0.07597621	-0.00525699	0.24980631	0.2446756
-0.40551217	-0.30407957	-0.13892575	0.04232061	0.10490774	0.14491331
-0.3249819	-0.42345507	-0.16704476	-0.05702717	0.23716445	0.24621087
1.22355057	0.64802323	0.8015674	-0.62443732	-0.54829595	0.15643979

Word Finding Subjects

DL vs. WF	DL vs. EO	DL vs. EC	WF vs. EO	WF vs. EC	EO vs. EC
-0.81461773	0.19233056	-0.75268235	1.03140067	0.07934742	-0.97806779
-1.12608935	-0.91829947	-0.79032885	0.31450686	0.49166624	0.07205494
-1.06482194	0.20589801	-0.80537106	1.28359134	0.29541465	-1.05157281
-0.97343363	-0.71202557	-0.64666043	0.30438232	0.44892595	0.06363267
0.03995363	-0.50925519	-0.24269429	-0.41227716	-0.20045957	0.22225459
-0.15746285	-0.68483341	-0.5125057	-0.36219216	-0.2798572	0.0844074
-1.08801049	-1.08269665	-1.00073482	0.19651473	0.13236103	-0.00978842
1.26466532	-0.40049906	0.87895518	-1.50848913	-0.41487932	1.14631996
0.2634058	-0.3864594	0.16038507	-0.65782494	-0.18482203	0.51424421
-0.94190334	-0.82229856	-0.95513389	0.27283012	0.00757822	-0.22932529

Eyes Open Subjects

DL vs. WF	DL vs. EO	DL vs. EC	WF vs. EO	WF vs. EC	EO vs. EC
-0.01024492	-0.22681234	-0.30795531	-0.11003162	-0.29253371	-0.17203594
0.38383151	-0.01749135	0.36825377	-0.44752399	-0.03544168	0.43600911
-0.80611644	0.00455575	-0.20545092	0.69213555	0.55399083	-0.18156525
-0.89646882	0.26333641	-0.70124851	1.15708151	0.20332608	-0.9996422
-0.74726867	-0.53058003	-0.49068665	0.20183559	0.13639825	0.06942238
-0.19250447	-0.23753957	-0.01569117	-0.06016934	0.20459853	0.22803452
0.51863263	0.13220302	0.6401859	-0.48352538	0.12289532	0.58335794
0.65622056	-0.44758602	-0.09701106	-0.86699066	-0.46487961	0.23641931
-0.4652939	-0.43803487	-0.41726457	0.10042181	0.10623288	-0.02689236
-0.29112424	-0.44688217	-0.13400429	-0.1473896	0.19788072	0.29857849
-0.47466987	-0.44151367	-0.33808692	0.08289878	0.20977871	0.0847115

Eyes Closed Subjects

DL vs. WF	DL vs. EO	DL vs. EC	WF vs. EO	WF vs. EC	EO vs. EC
-0.07171376	-0.24648963	0.10241652	-0.19056749	0.22495196	0.3871293
0.38804833	-0.00866984	0.44574127	-0.46389415	0.03923021	0.5145132
-1.43745215	-0.71001145	-0.99379605	0.80603445	0.59820775	-0.31442793
-1.30849452	-1.1826236	-1.0691743	0.31819239	0.42659228	0.0059935
1.13819433	1.2793813	1.2787721	-0.13164559	-0.01855573	0.13753579
0.2804833	-0.38637858	-0.31346992	-0.49554344	-0.40421255	0.05643391
0.9927695	0.19131219	0.60026603	-0.79847823	-0.37083664	0.41438896
-0.8496114	-0.42421648	-0.65300776	0.49894928	0.32751618	-0.26336412
1.34715231	1.47347293	1.45709856	-0.11992308	-0.10788032	0.10657605
-0.68136626	-0.07652033	-0.27322878	0.49485488	0.27592779	-0.1766731
-1.5458179	-1.35323453	-1.197498	0.39604375	0.54915644	0.05644208

Table 10.5: Reliabilities and Efficiencies, Test Data Set

Task	Reliability	Efficiency
Dot Localization	23/32	0
Word Finding	32/33	0
Eyes Open	27/32	1/11
Eyes Closed	-	-

The direct rules to identify tasks fail mainly because all efficiencies fall to around 0.

10.4 Ruling Out Conditions

10.4.1 Rules

The conditions to rule out tasks (Table 10.6) are defined by straightforward numerical values and all reliabilities are over 0.8. Efficiencies are greater than 0.5 for active tasks but less than 0.5 for the passive ones.

Table 10.6: Ruling Out Conditions, Reliabilities and Efficiencies, Rule Data Set

Task	Rule Out Task If:	Reliability	Efficiency
Dot Localization	DL vs. EC > 0	9/11	19/34
Word Finding	DL vs. WF < 0	8/10	19/35
Eyes Open	WF vs. EO < -0.6	11/12	10/33
Eyes Closed	EO vs. EC < -0.2	11/12	12/33

10.4.2 Testing

Tests of ruling out conditions fail (Table 10.7). Dot localization, word finding and eyes closed are erroneously ruled out more often than they are correctly ruled out. Eyes open does not fare much better.

Table 10.7: Reliabilities and Efficiencies, Test Data Set

Task	Reliability	Efficiency
Dot Localization	7/11	9/32
Word Finding	3/10	20/33
Eyes Open	10/11	6/32
Eyes Closed	9/11	5/32

10.5 Discussion

Direct rules and ruling out conditions both fail their tests. Such poor discrimination was predicted by the low statistical distances between tasks. Using highest slope to find task channel pairs for use in discrimination would seem to be inferior to highest correlation.

Chapter 11

Conclusions

While discrimination using phase alone proved fruitless, methods using the correlation, slope and direction of phase/frequency of pairs of channels receiving similar but lagged signals showed some promise. First, since linear relationships and not just noise existed, there was a chance that relationships would differ between tasks. Secondly, it was possible to develop classification rules using phase/frequency direction, slope and correlation. However, rules did not hold up well upon testing.

11.1 Evaluation of Phase/Frequency

Tables 11.1 and 11.2 provide performance summaries for all methods used to develop classification rules. Results are, at best, marginally successful.

(i) Classification rules which meet 0.8 reliability can be devised, but efficiencies of 0.5 can rarely be achieved. Not even the 0.8 reliabilities hold under testing.

(ii) Only one method, highest correlation task pairs correlation/slope vectors, provides rules which can discriminate among all four tasks at once. In testing, the four-way classification suffers from fluctuating reliabilities and falling efficiencies.

However, the rules pass in that the probability of Type I errors does not exceed the probability of correct classifications. Two other methods, highest correlation subject pairs and highest correlation task pairs with PCA, can only discriminate among three tasks with the same dubiously successful performance.

(iii) Phase/frequency does not contain enough task oriented differences to allow any direct task identification and ruling out conditions are the best that can be found.

Table 11.1: Summary of Categorical Ruling Out Conditions

		Rule			Test		
Method	Task	Rel.	Eff.	Sum	Rel.	Eff.	Sum
Highest Correlation Subject Pairs	DL	0.82	0.62	1.44	0.27	0.81	1.08
	WF	-	-	-	-	-	-
	EO	1	0.30	1.30	0.73	0.38	1.11
	EC	0.92	0.27	1.19	0.82	0.41	1.23
Highest Correlation Task Pairs with PCA	DL	1	0.21	1.21	0.64	0.5	1.14
	WF	1	0.2	1.2	1	0.15	1.13
	EO	-	-	-	-	-	-
	EC	1	0.15	1.15	0.91	0.13	1.04
Highest Correlation Task Pairs without PCA	DL	1	0.24	1.24	1	0.25	1.25
	WF	0.92	0.2	1.12	0.8	0.18	0.98
	EO	1	0.64	1.64	0.73	0.22	0.95
	EC	0.92	0.64	1.55	0.91	0.13	1.04

Table 11.2: Summary of Numerical Direct Rules and Ruling Out Conditions

		Direct Identification						Ruling Out Conditions					
		Rule			Test			Rule			Test		
Method	Task	Rel.	Eff.	Sum	Rel.	Eff.	Sum	Rel.	Eff.	Sum	Rel.	Eff.	Sum
Highest Correlation Task Pairs Correlation/Slope Vectors	DL	0.82	0.55	1.37	0.97	0	0.97	0.91	0.65	1.56	0.27	0.97	1.24
	WF	-	-	-	-	-	-	0.90	0.54	1.44	0.70	0.36	1.06
	EO	0.82	0.33	1.15	0.90	0.09	0.99	0.83	0.27	1.10	1	0.03	1.03
	EC	0.85	0.67	1.52	0.28	0.91	1.19	0.83	0.39	1.22	0.91	0.25	1.16
Highest Slope Task Pairs Correlation/Slope Vectors	DL	0.91	0.45	1.36	0.72	0	0.72	0.82	0.56	1.38	0.64	0.28	0.92
	WF	0.91	0.40	1.31	0.97	0	0.97	0.8	0.54	1.34	0.3	0.61	0.91
	EO	0.88	0.58	1.46	0.84	0.09	0.93	0.92	0.30	1.22	0.91	0.19	1.10
	EC	-	-	-	-	-	-	0.92	0.36	1.28	0.82	0.16	0.98

11.2 Problems with Phase/Frequency

In developing rules that are direct, reliable and efficient, most of the weaknesses in phase/frequency seem to stem from two problems, lack of distance between tasks and overlapping of tasks.

Distance refers to how much correlation, slope and direction of phase/frequency for channel pairs receiving similar but lagged signals differ between tasks. The statistical distances in the correlation/slope vectors (Sections 9.2 and 10.2) show that tasks are less than two standard deviations apart. It is evident that the between-task to within-task variation ratio is too low.

Overlap between tasks occurs because channel pair phase/frequency correlation, slope and direction behave the same way for more than one task. This can be seen in the overlapping of highest correlation subject channel pairs and highest correlation task pairs plot categorizations. The between-task variation in absolute terms is low.

In addition, the amount of data processing needed to use phase/frequency might be prohibitive. It involves computing channel pair covariances, smoothing, calculating phases and storing them as phase arrays, reducing data to significant channel pairs, and processing the phase/frequency properties of channel pairs to find task dependent differences. Doing this on enough data sets to develop classification rules which are trustworthy and then repeating the process to use the rules seems impractical.

11.3 Summary

The goals of the thesis have not been met.

There is some information in phase/frequency for some groupwise discrimination, but there is not enough to allow rules to be either simple or direct and not enough to discriminate among four tasks reliably and efficiently.

From this study and with this data set, it seems that neither phase nor phase/frequency will improve upon methods of discrimination which have used power spectra and cross-spectra.

Bibliography

[1] Brillinger, D.R. (1981). Time Series: Data Analysis and Theory. University of California, Berkeley, 488 pp.

[2] Brown, J.W. and Churchill, R.V. (1996). Complex Variables and Applications. McGraw-Hill, 386 pp.

[3] Freund, J.E. (1992). Mathematical Statistics (5th ed.). Prentice Hall, 658 pp.

[4] Gonzalez, R.C. and Wintz, P. (1977). Digital Image Processing. Addison-Wesley Publishing Company, Advanced Book Program/World Science Division, Reading, Massachusetts, 553 pp.

[5] Hannan, E.J. (1983). Signal estimation. In: Time Series in the Frequency Domain, Handbook of Statistics, Vol. 3, D.R. Brillinger and P.R. Krishnaiah, (eds.), Elsevier Science Publishers, pp. 111-123.

[6] Koles, Z.J., Professor of Electrical and Computer Engineering and Biomedical Engineering, University of Alberta, Edmonton, pers. comm.

[7] Lind, J.C., Koles, Z.J., Flor-Henry, P., and Soong, C.K. (1997). Frequency domain discriminant analysis of the electroencephalogram. Psychophysiology, 34, 358-364.

[8] Lind, J.C., Flor-Henry, P., and Koles, Z.J. (1999). Discriminant analysis and equivalent source localization of the EEG related to cognitive functions. Brain Topography, 11, 265-278.

[9] Lind, J.C., Quantitative Psychologist, Clinical Diagnostics and Research Centre, Alberta Hospital, Edmonton, pers. comm.

[10] Morrison, D.F. (1990). Multivariate Statistical Methods (3rd ed.). New York, McGraw-Hill, 415 pp.

[11] Pfurtscheller, G., Auer, L.M., and Kopruner, V. (1984). Quantitative EEG follow-up study after extracranial-intracranial bypass operation for cerebral ischemia. In: Brain Ischemia: Quantitative EEG and Imaging Techniques, Progress in Brain Research, Vol. 62, G. Pfurtscheller, E.J. Jonkman, and F.H. Lopes da Silva, (eds.), Elsevier Science Publishers, pp. 121-143.

[12] Priestly, M.B. (1981). Spectral Analysis and Time Series. New York, Academic Press, 2 Vols.

[13] Rao, C.R. (1973). Linear Statistical Inference and Its Applications (2nd ed.). New York, Wiley, 625 pp.

[14] Rawlings, J.O. (1988). Applied Regression Analysis: A Research Tool. Wadsworth & Brooks/Cole, Statistics/Probability Series, Pacific Grove, California, 553 pp.

[15] Sharbrough, F., Chatrian, G., Lesser, R.P., Luders, H., Nuwer, M., and Picton, T.W. (1990). Guidelines for standard electrode position nomenclature. American Electroencephalographic Society. Bloomfield, CT.

[16] Shumway, R.H. (1988). Applied Statistical Time Series Analysis. Division of Statistics, University of California, Davis. Prentice Hall, Englewood Cliffs, New Jersey, 379 pp.

[17] Van Huffelen, A.C., Poortvliet, D.C.J., and van der Wulp, C.J.M. (1984). Quantitative electroencephalography in cerebral ischemia. Detection of abnormalities in "normal" EEGs. In: Brain Ischemia: Quantitative EEG and Imaging Techniques, Progress in Brain Research, Vol. 62, G. Pfurtscheller, E.J., Jonkman, and F.H. Lopes da Silva, (eds.), Elsevier Science Publishers, pp. 3-28.

Appendix A

Channel Number Nomenclature

Channel Number	Site Name	Channel Number	Site Name
1	AF_3	23	AF_z
2	AF_4	24	$*F_z$
3	FC_5	25	FC_z
4	FC_6	26	$*C_z$
5	FC_1	27	CP_z
6	FC_2	28	$*P_z$
7	$*C_3$	29	PO_z
8	$*C_4$	30	$*Fp_1$
9	C_1	31	$*Fp_2$
10	C_2	32	$*F_7$
11	TP_7	33	$*F_8$
12	TP_8	34	$*F_3$
13	CP_5	35	$*F_4$
14	CP_6	36	FT_7
15	CP_1	37	FT_8
16	CP_2	38	$*T_7$=OLD T_3
17	$*P_3$	39	$*T_8$=OLD T_4
18	$*P_4$	40	$*P_7$=OLD T_5
19	P_1	41	$*P_8$=OLD T_6
20	P_2	42	$*O_1$
21	PO_3	43	$*O_2$
22	PO_4		