ON THE STUDY OF MULTI-CHANNEL EEG: LOSSLESS COMPRESSION, SIGNAL MODELING AND CLASSIFICATION

by

YODCHANAN WONGSAWAT

Presented to the Faculty of the Graduate School of

The University of Texas at Arlington in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT ARLINGTON

December 2007

ACKNOWLEDGEMENTS

I would like to thank my supervising professor Dr. Soontorn Oraintara for constantly motivating and encouraging me, and also for his invaluable advice during the course of my doctoral studies. I am highly grateful to Dr. K. R. Rao for providing me useful guidance on my research. I would like to thank Dr. Michael T. Manry for his comments on my classification research. I wish to thank my academic advisors Dr. Zhou Wang, Dr. Jean Gao, and Dr. Qilian Liang for their interests in my research and for taking time to serve in my dissertation committee.

I would like to extend my appreciation to Dr. Toshihisa Tanaka, Tokyo University of Agriculture and Technology and RIKEN Brain Science Institute, for providing the multi-channel EEG data for the experiments. I would like to thank Kianoush Nazarpour and Dr. Saeid Sanei, the school of engineering, Prof. Edward Wilding, the school psychology, Cardiff University, for kindly discussions and providing the dataset for the artifact removal experiment. Furthermore, I would also like to thank Dr. Nuri Firat Ince and Dr. Ahmed Tewfik, Department of Electrical and Computer Engineering, University of Minnesota, for the discussions and the LDB Toolbox.

My special thanks go to my MSP labmates, i.e. Yothin Rakvongthai, An Vo, Nha Nguyen, Dr. Truong T. Nguyen, and Dr. Yilong Liu for their helpful comments on my research. Many thanks also go to my friends at UTA, Chivalai Temiyasathit, Dr. Prattana Punnakitikashem, Att Kruafak, Dr. Chai Chompoo-Inwai, Thitipong Wattanachewanopakorn, Wikrom Prombutr, and Titi Saksornchai for making my life here pleasurable.

My deep gratitude goes to my family for inspiring and supporting me toward my doctoral degree. Finally, I would like to particularly thank Nantakan Silseweekul for comforting me during the hard time.

November 15, 2007

ABSTRACT

ON THE STUDY OF MULTI-CHANNEL EEG: LOSSLESS COMPRESSION, SIGNAL MODELING AND CLASSIFICATION

Publication No. _____

YODCHANAN WONGSAWAT, Ph.D.

The University of Texas at Arlington, 2007

Supervising Professor: Soontorn Oraintara

Recently, electroencephalogram (EEG) has become necessary for diagnosis, telemedicine, and brain computer interface (BCI). This thesis investigates three signal processing tools which can explore multi-channel structures of the EEG for different applications such as lossless compression, artifact removal, and classification.

First, this thesis presents a method for approximating the Karhunen-Loeve transform (KLT) of multi-channel EEG signals. The proposed transform is further parameterized by lifting factorization by which the coefficients' dynamic range is controlled through a proposed pivoting scheme, rendering a reversible structure under quantization of coefficients called IntSKLT. A lossless coder for multi-channel EEG signals which exploits inter-channel correlation among the EEG channels by the IntSKLT is also presented. Simulation results show that the coding performance of the proposed coder is improved by approximately 10% over the benchmark lossless coders. Furthermore, compared with directly using the reversible structure of the KLT (IntKLT), the degradation in coding

performance using the IntSKLT is approximately 3% while the computational complexity is reduced by more than 60%.

Second, in order to avoid high computational load on calculating the parallel factor analysis (PARAFAC)-based space-time-frequency model of a multi-channel EEG signal, this thesis presents three reduced complexity space-time-frequency models. The models are developed by dividing the selected contents into segments followed by applying the PARAFAC. By carefully selecting the numbers of segments, signatures extracted from the conventional space-time-frequency model can be approximated by those from the proposed models with the computational complexity reduced by more than 50%. Simulation results show that the proposed models can efficiently extract eyeblink artifacts from background EEG. Furthermore, classification accuracies when employing the proposed models to the BCI application are also comparable with the conventional model.

Finally, this thesis presents a feature extraction scheme called multi-channel flexible local discriminant bases (MF-LDB) for left/right imagery classification of a multi-channel EEG. The MF-LDB is obtained by calculating the local cosine packets (LCP) of the decided channel over nonuniform segments. The proposed method combines information from neighboring channels based on hard and soft decisions. Simulation results show that the proposed feature extraction scheme can improve the classification accuracies of the left/right imagery signals by more than 3%. By applying the minimum variance distortionless response (MVDR) to find the spectra over nonoverlapping segments of each EEG channel, a nonredundant time-frequency transform called local MVDR packets transform which can provide highly selective frequency responses is also presented with approximately 4% improvement in classification accuracy.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1 Problem Statement

Electroencephalogram (EEG) has been widely used for clinical analysis (e.g. to diagnose the disease and to assess the effectiveness of the treatment via the brain functions [4]) and also for the brain computer interface (BCI), i.e. using brain signal instead of hand. Generally, EEGs are measured by placing many electrodes onto human scalp, resulting in a multi-channel EEG signal. Each channel represents an EEG signal at different positions on the scalp. Since the human scalp has a limited area, some neighboring channels (electrodes) of EEG can be highly correlated. However, when some particular activities occur, e.g. eyeblink, thinking, movement, some neighboring channels can be quite different. This thesis aims to develop signal processing tools for exploiting the usefulness of these multi-channel structures of EEG to solve two major problems, which are

1. how to losslessly compress the huge amount of EEG data, and
2. how to efficiently model an EEG signal in order to extract valuable information for the purpose of analysis, e.g. artifact removal and classification.

1.2 Contributions of the Thesis

Three major contributions of this thesis can be summarized as follows:

1. In order to solve the first problem in 1.1 an investigation on approximation of the Kahunen-Loeve transform (KLT) used in order to decorrelate a multi-channel EEG for lossless compression is presented. Inter-channel correlation of the multi-channel EEG has

been exploited to losslessly compress a multi-channel EEG via the sub-optimal KLT. This transform is obtained by subdividing the data into segments and then iteratively applying a local KLT to each segment of the data. The reversible approximation of the sub-optimal KLT called IntSKLT is developed via the lifting-type factorization. A pivoting scheme for selecting the permutation matrices which leads to less rounding errors for the factorization is also proposed. Since the sub-optimal KLT consists of many smaller-sized KLT matrices, factorizing these matrices would simplify the reversible approximation process. The coding algorithm using the IntSKLT for decorrelating the inter-channel correlation, and using the autoregressive model of order six and reversible discrete cosine transform of type IV (IntDCT-IV) for temporal decorrelation, is also introduced. This research can be useful for telemedicine since sometimes the information has to be losslessly transmitted.

2. In order to solve the second problem on modeling a multi-channel EEG in 1.1, reduced complexity space-time-frequency models are proposed. This topic focuses on reducing the computational complexity of this 3-way PARAFAC model by the proposed 4- and 5-way PARAFAC models. Basically, the complexity is reduced by segmenting the selected domains (space and time) to form the 4- or 5-dimemsional array data. Then the 4- or 5-way PARAFAC model is applied to this new data in order to simultaneously extract the important features contained in all space, time, frequency, and segment domains. By using our derived equations with a constraint on the length of segments, the features of the conventional 3-way model can be approximated by those of the reduced complexity models. The proposed models are useful in many applications, such as artifact removal and classification. The application of one of the reduced complexity model, i.e. space-time-frequency-time/segment (STF-TS) model, on removing of the eyeblink artifact is illustrated. By employing the space signature of the STF-TS model as the

estimated steering vector for the beamformer, the eyeblink artifact contaminated in a multi-channel EEG can be efficiently removed.

3. To improve the classification accuracy for the left/right imagery EEG, the feature extraction scheme called multi-channel flexible local discriminant bases has been developed. In this research, the information from multiple channels is exploited to design the data dependent features via the local cosine packets transform (LCP) (also known as dual of the wavelet packets transform). Specifically, the nonuniform segments designed from the LCP coefficients of multiple channels are designed to optimally discriminate between the left and right imagery EEGs. The high classification accuracy can be obtained by employing the proposed features. To further improve the classification accuracy, we also develop a data dependent time-frequency transform similar to the local cosine packets by calculating the spectra of each non-overlapping window through the signal. These spectra are computed so that the variances of the filtered signals in each window are minimized. The possible applications of this issue are such as controlling the wheelchairs, mouse, or remote by using brain signals, e.g. EEG.

1.3 Organization of the Thesis

This thesis begins with the background (Chapter 2) on a multi-channel EEG and useful signal processing tools which will be used for the rest of the thesis. Chapter 3 presents a sub-optimal KLT and its application on lossless multi-channel EEG compression. Low complexity models for a multi-channel EEG are presented in Chapter 4 together with some of their applications on artifact removal and left/right imagery EEG classification. In Chapter 5, we introduce a feature extraction scheme that takes into account multiple channels of EEG for improving the classification accuracy of the left/right imagery EEG classification. Finally, Chapter 6 concludes this thesis.

CHAPTER 2

BACKGROUND

This chapter aims to provide the background on a multi-channel EEG and signal processing tools that will be used in this thesis. This includes the background on Karhunen-Loeve transform (KLT), matrix factorizations, space-time-frequency modeling, local cosine packets (LCP), and minimum variance distortionless response (MVDR).

2.1 Multi-channel EEG

EEG is one type of brain signals which is measured by placing electrodes, i.e. small disks made of silver-silver chloride (Ag-AgCl), on the human scalp. The resulting EEG data obtained by measuring EEGs using more than one electrode at a time is called *multi-channel EEG* (each electrode is sometimes called channel).

2.1.1 Electrode Positioning

Electrodes are usually placed on the human scalp according to the recommendation of the International Federation of Societies for Electroencephalography and Clinical Neurophysiology called 10-20 system (as depicted in Fig. 2.1). The 10-20 system is designed to avoid the eyeball placement. By using specific anatomic landmarks from which the measurement would be made, 10 or 20% of that specified distance is used as the position of the electrodes.

2.1.2 Brain Rhythms

There are five main rhythms (frequency bands) which are used to categorize EEG signals into five different types of waves, i.e. delta, theta, alpha, beta, and gamma waves (Fig. 2.2):

1. **Delta (δ) waves** lie within the frequency range of 0.5-4 Hz. These waves are generally associated with deep sleep.
2. **Theta (θ) waves** lie within the frequency range of 4-8 Hz. These waves appear as consciousness slips towards drowsiness.
3. **Alpha (α) waves** lie within the frequency range of 8-13 Hz. These waves usually indicate relaxed awareness without any concentration. Alpha-range signal which is seen over the sensorimortor cortex is called Mu (μ) rhythms. This μ rhythm is very useful to discriminate the imagery signal in the brain computer interface (BCI).
4. **Beta (β) waves** lie within the frequency range of 13-30 Hz. These waves are associated with active thinking.
5. **Gamma (γ) waves** have the frequency range above 30 Hz. These waves have low amplitude but high frequency and can be used to confirm some brain diseases. The occurrence of these waves is rare.

More detail on a fundamental of EEG can be found in [1].

2.2 Karhunen-Loeve Transform (KLT)

Karhunen-Loeve transform (KLT) is the transform that maps a real random vector $\mathbf{x} = (x_0, x_1, \ldots, x_{N-1})^T$ to a random vector $\mathbf{y} = (y_0, y_1, \ldots, y_{N-1})^T$ such that $\mathbf{y}$ is completely decorrelated. In particular, if $\mathbf{\Phi}^T$ is the matrix representation of the KLT, then $\mathbf{y} = \mathbf{\Phi}^T\mathbf{x}$ and

$$E\left\{\mathbf{y}\mathbf{y}^T\right\} = E\left\{\mathbf{\Phi}^T\mathbf{x}\mathbf{x}^T\mathbf{\Phi}\right\} = \mathbf{\Phi}^T E\left\{\mathbf{x}\mathbf{x}^T\right\}\mathbf{\Phi} = \mathbf{\Phi}^T\mathbf{R}_{xx}\mathbf{\Phi} = \mathbf{\Lambda},$$

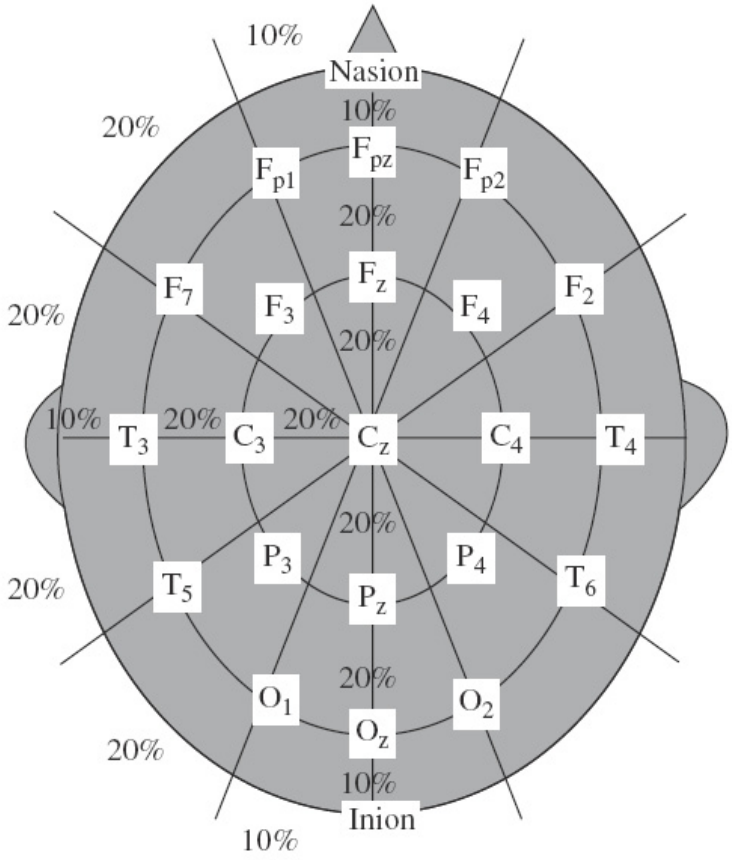

Figure 2.1. Conventional 10-20 electrode positions for measuring a multi-channel EEG signal [1].

where the columns of $\mathbf{\Phi}$ are the normalized eigenvectors of $\mathbf{R}_x = E\left\{\mathbf{x}\mathbf{x}^T\right\}$, $\mathbf{\Lambda}$ is the diagonal matrix of the eigenvalues with respect to $\mathbf{\Phi}$ and T denotes the transpose operation.

Case of Markov-1 Model

By assuming the input signal as the first order stationary Markov process, the analytical solution for the KLT matrix can be solved. The correlation between the i-th and j-th channels is given by:

$$[\mathbf{R}_{xx}]_{ij} = E\{x_i x_j\} = \rho^{|i-j|}, \quad i, j = 0,\ 1, \cdots,\ N-1, \tag{2.1}$$

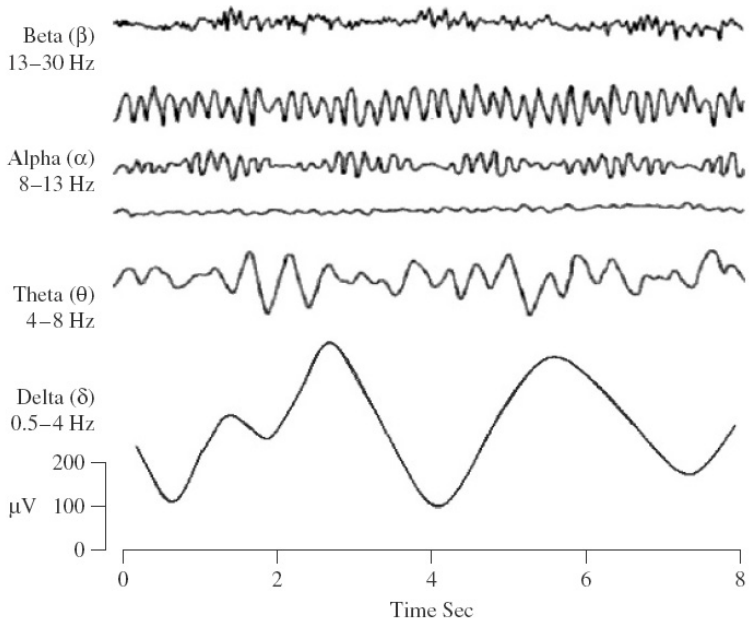

Figure 2.2. The rhythms of EEG signals [1].

where ρ is the adjacent correlation coefficient such that $0 < \rho < 1$. The analytical eigenvalues and their corresponding eigenvectors for this special class of inputs are given in [5].

2.3 Matrix Factorization

There are many ways to parameterize a non-singular matrix such as **QR** factorization, singular value decomposition (SVD) and **LU** factorization. In this section, for the purpose of constructing a reversible transform, a modification of the **LU** factorization is summarized below. The idea is that a non-singular matrix **A** is factorized so that both **L** and **U** matrices have the diagonal elements equal to 1 or -1. For detailed discussion, the reader is referred to [6]:

Theorem 1 [6]: Given a non-singular diagonal matrix $\mathbf{D}_R = \text{diag}(1, 1, \ldots, 1, e^{i\theta})$, the matrix **A** has a factorization of $\mathbf{A} = \mathbf{PLD}_R\mathbf{US}_0$ if and only if det $(\mathbf{P}^T\mathbf{A})$ = det

$(\mathbf{D}_R) \neq 0$, where $\mathbf{P}$ is a permutation matrix, $\mathbf{L}$, $\mathbf{U}$, $\mathbf{S}_0$ are a lower triangular matrix, an upper triangular matrix, a single row marix, respectively, and their diagonal elements are $1, -1, i$, or $-i$.

In this thesis, we are interested in this form of matrix factorization when $e^{i\theta}$ in $\mathbf{D}_R$ is 1 or -1, i.e. the matrix factorization can be reduced to the form of $\mathbf{A} = \mathbf{PLUS}_0$, where the diagonal elements of $\mathbf{L}$ and $\mathbf{U}$ are 1 or -1. The algorithm for constructing the matrix factorization in *Theorem 1* can be summarized as follows:

1) Suppose that $\mathbf{A}$ is an $N \times N$ non-singular matrix. There exists a permutation matrix $\mathbf{P}_1$ for row exchanging such that

$$\mathbf{A}^{(1)} = \mathbf{P}_1\mathbf{A} = \begin{bmatrix} \hat{a}_{1,1}^{(1)} & \hat{a}_{1,2}^{(1)} & \cdots & \hat{a}_{1,N}^{(1)} \\ \hat{a}_{2,1}^{(1)} & \hat{a}_{2,2}^{(1)} & \cdots & \hat{a}_{2,N}^{(1)} \\ \cdots & \cdots & \cdots & \cdots \\ \hat{a}_{N,1}^{(1)} & \hat{a}_{N,2}^{(1)} & \cdots & \hat{a}_{N,N}^{(1)} \end{bmatrix},$$

with $\hat{a}_{1,N}^{(1)} \neq 0$.

2) According to 1), there exists a matrix $\mathbf{S}_{0,1} = \mathbf{I} - s_1 e_N e_1^T$ that makes the pivot on the first row of $\mathbf{A}^{(1)}$ equal to 1, where $s_1 = (\hat{a}_{1,1}^{(1)} - 1)/\hat{a}_{1,N}^{(1)}$ and e_j is the column vector of which the j-th element equals to 1 and the rest equal to zero, i.e.

$$\mathbf{P}_1\mathbf{A}\mathbf{S}_{0,1} = \begin{bmatrix} 1 & \hat{a}_{1,2}^{(1)} & \cdots & \hat{a}_{1,N}^{(1)} \\ \hat{a}_{2,1}^{(1)} - s_1\hat{a}_{2,N}^{(1)} & \hat{a}_{2,2}^{(1)} & \cdots & \hat{a}_{2,N}^{(1)} \\ \cdots & \cdots & \cdots & \cdots \\ \hat{a}_{N,1}^{(1)} - s_1\hat{a}_{N,N}^{(1)} & \hat{a}_{N,2}^{(1)} & \cdots & \hat{a}_{N,N}^{(1)} \end{bmatrix}.$$

3) Apply the elimination matrix $\mathbf{L}_1$ in order to achieve the forward elimination of the first column,

$$\mathbf{L}_1\mathbf{P}_1\mathbf{A}\mathbf{S}_{0,1} = \begin{bmatrix} 1 & a_{1,2}^{(1)} & \cdots & a_{1,N}^{(1)} \\ 0 & a_{2,2}^{(1)} & \cdots & a_{2,N}^{(1)} \\ \cdots & \cdots & \cdots & \cdots \\ 0 & a_{N,2}^{(1)} & \cdots & a_{N,N}^{(1)} \end{bmatrix}.$$

4) Continue the algorithm for $N-1$ iterations. Finally, we obtain

$$\mathbf{L}_{N-1}\mathbf{P}_{N-1}\ldots\mathbf{L}_1\mathbf{P}_1\mathbf{A}\mathbf{S}_{0,1}\ldots\mathbf{S}_{0,N-1} = \mathbf{U}$$
$$= \begin{bmatrix} 1 & a_{1,2}^{(N-1)} & \cdots & a_{1,N}^{(N-1)} \\ 0 & 1 & \cdots & a_{2,N}^{(N-1)} \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & a_{N,N}^{(N-1)} \end{bmatrix},$$

where $a_{N,N}^{(N-1)} = \pm 1$.

5) Thus, we can obtain $\mathbf{L}^{-1}\mathbf{P}^T\mathbf{A}\mathbf{S}_0^{-1} = \mathbf{U}$, i.e. $\mathbf{A} = \mathbf{P}\mathbf{L}\mathbf{U}\mathbf{S}_0$, where

$$\mathbf{S}_0^{-1} = \mathbf{S}_{0,1}\ldots\mathbf{S}_{0,N} = \begin{bmatrix} 1 & & 0 & \\ & I & & \\ 0 & & 1 & \\ -s_1 & \cdots & -s_{N-1} & 1 \end{bmatrix},$$

$$\mathbf{L}^{-1} = \mathbf{L}_{N-1}(\mathbf{P}_{N-1}\mathbf{L}_{N-2}\mathbf{P}_{N-1}^T)\ldots(\mathbf{P}_{N-1}\ldots\mathbf{P}_2\mathbf{L}_1\mathbf{P}_2^T\ldots\mathbf{P}_{N-1}^T), \quad \text{and}$$

$$\mathbf{P}^T = \mathbf{P}_{N-1}\ldots\mathbf{P}_2\mathbf{P}_1.$$

According to [6], we can further factorize the matrix $\mathbf{L}\mathbf{U}$ (in $\mathbf{A} = \mathbf{P}\mathbf{L}\mathbf{U}\mathbf{S}_0$) into N single-row elementary reversible matrices (SERMs), i.e. the matrices that the off-diagonal elements are all zero except for one row which contains the non-zero elements,

and the diagonal elements are ± 1, which will result in reducing of error from the rounding operations. The resulting factorization is now denoted as

$$\mathbf{A} = \mathbf{P}_L \mathbf{S}_N \mathbf{S}_{N-1} \cdots \mathbf{S}_1 \mathbf{S}_0, \tag{2.2}$$

where $\mathbf{P}_L$ is a permutation matrix, and $\mathbf{S}_i$ are SERMs for $i = 0, 1, \ldots, N$. It should be noted that the diagonal elements of $\mathbf{S}_i$ are ± 1.

Since, the choice of $\mathbf{P}_L$ is not unique and can affect the dynamic range of the non-zero elements in $\mathbf{S}_i$, choosing the right $\mathbf{P}_L$ is very important. However, if the dimension of $\mathbf{A}$ is large, this task can be very difficult.

2.4 Space-Time-Frequency Model

This section reviews some backgrounds on the STF model. Each channel of the 1-D time-domain signal is first transformed to reveal its 2-D time-frequency representation. By stacking all the 2-D time-frequency arrays from all the channels, we form a 3-D array in space-time-frequency domain. Then, the 3-way parallel factor analysis (PARAFAC) [7] is further applied to this 3-D array in order to decompose the data into its fundamental components yielding the STF model.

2.4.1 Time-Frequency Transform

In order to map a 1-D signal in time domain to a 2-D signal in time-frequency domain, time-frequency transform is employed. Time-frequency modeling is known to be practical for the analysis of 1-D nonstationary signals e.g. EEG [8], [9]. There are two main methods to achieve this goal, i.e. to simultaneously localize signals in both time and frequency domains, the Cohen's class (translate signal in time and frequency) and the affine class (translate signal for time resolution and scale the signal for frequency resolution). Since the affine class yields nonuniform nature of time-frequency signal

components, it is more suitable for EEG [10]. An EEG signal, $s(t)$, can be efficiently decomposed into the affine class time-frequency atoms by convolving with the complex Morlet wavelet basis (filter), $w(f,t)$, as

$$\acute{y}(f,t) = |w(f,t) * s(t)|^2 . \tag{2.3}$$

By stacking $\acute{y}(f,t)$ of all channels, a 3-D array can be formulated as $\acute{y}(n,f,t)$, where n is the channel index [10].

2.4.2 Space-Time-Frequency Model (STF Model)

In order to decompose a 3-D array signal into space, time and frequency domains, the 3-way PARAFAC is applied to the 3-D array signal, $\acute{y}(n,f,t)$ (denoted in array form as $\underline{\acute{\mathbf{Y}}}$) resulting in the STF model, which can be formulated as

$$\underline{\acute{\mathbf{Y}}}_{N\times F\times T} = h(\acute{\mathbf{A}}, \acute{\mathbf{C}}, \acute{\mathbf{D}}) + \underline{\acute{\mathbf{E}}}_{N\times F\times T}, \tag{2.4}$$

where the 3-way PARAFAC model, i.e. the STF model, is

$$h(\acute{\mathbf{A}}, \acute{\mathbf{C}}, \acute{\mathbf{D}}) = \sum_{m=1}^{M} \acute{a}(n,m)\acute{c}(f,m)\acute{d}(t,m),$$

and $\underline{\acute{\mathbf{E}}}$ is a 3-D array residual of the model. Each column of $\acute{\mathbf{A}}_{N\times M}$ denotes a space signature of the m-th component where its matrix elements are denoted as $\acute{a}(n,m)$, n is the channel index ranging from 1 to N, m is the component index ranging from 1 to M, and M is the number of components. Each column of $\acute{\mathbf{C}}_{F\times M}$ denotes the frequency signature where its matrix elements are denoted as $\acute{c}(f,m)$ and f is the frequency index ranging from 1 to F. Each column of $\acute{\mathbf{D}}_{T\times M}$ denotes the time signature where its matrix elements are denoted as $\acute{d}(t,m)$, and t is the time index ranging from 1 to T. It is noted that a suggested number of components M should be the one that maximizes the core consistency diagnostic (CORCONDIA) value which in [7] is known as an efficient model

validation criteria. The parameters $\acute{\mathbf{A}}$, $\acute{\mathbf{C}}$, and $\acute{\mathbf{D}}$ can be estimated by using the alternate least square algorithm (ALS) [7] where the cost function is

$$\text{argmin}_{\acute{a},\acute{c},\acute{d}} \left\| \underline{\acute{\mathbf{Y}}} - \sum_{m=1}^{M} \acute{a}(n,m)\acute{c}(f,m)\acute{d}(t,m) \right\|.$$

Intuitively, the space signatures in $\acute{\mathbf{A}}$ obtained from this STF model represent the weighting parameters of the inter-channel correlation among time-frequency representations of each channel. Taking into account that this STF model needs to simultaneously process a 3-D array signal, hence, if at least one of its three dimensions, i.e. space, time or frequency, is large, the decomposition will be very complex and makes this elegant model infamous for real-world applications.

2.5 Local Cosine Packets (LCP)

Sometimes original data can be represented in a more meaningful fashion (according to some applications) by transforming it into another domain. In order to preserve all information with the minimum number of transform coefficients, the selected transform should be nonredundant and orthogonal. To achieve this, the LCP is used instead of the short time Fourier transform (STFT) (with the overlapping windows), which is redundant. The LCP is similar to the wavelet packets in the sense that it provides a time-frequency representation with uniform partitioning in frequency domain.

Let $x(t)$ be a signal in time domain, and $[a_j, a_{j+1}]_{j\in Z}$ be a set of partitions of $x(t)$ where the length of each partition is $l_j = a_{j+1} - a_j$. LCP coefficients can be calculated by

$$C_k^j = \left\langle x(t), \psi_k^j(t) \right\rangle = \int_{a_j-\gamma}^{a_{j+1}+\gamma} x(t)\psi_k^j(t)dt, \tag{2.5}$$

where

$$\psi_k^j(t) = w_j(t)\sqrt{\frac{2}{l_j}}\cos\left[\pi\left(k+\frac{1}{2}\right)\frac{(t-a_j)}{l_j}\right], j \in \text{ Z}, k \in \text{Z}^+,$$

and γ is overlapping part of the window $w_j(t)$ which is less than or equal to l_j. To preserve the orthogonality, the smooth window function $w_j(t)$ is constructed using the cutoff function $b(t)$ which satisfies the following conditions:

$$\left|b(t)^2\right| + \left|b(-t)^2\right| = 1 \quad \text{for} \quad t \in \Re,$$

$b(t) = 0$ if $t < -1$ and $b(t) = 1$ if $t > 1$. The reader is referred to [11] for more details on the LCP.

2.6 Minimum Variance Distortionless Response (MVDR)

The minimum variance distortionless response (MVDR) can be used to estimate the spectrum of the time-domain signal $y(t)$ by employing the Fourier matrix into the optimization process [12]. Specifically, let $y(t), t = 0, ..., \hat{T}-1$, be the signal obtained from passing a discrete time signal $x(t)$ to the length p bandpass filter $\mathbf{a} = [a(0), ..., a(p-1)]^T$ with center frequency at ω, where $\hat{T}$ is the length of the signal. Let $E[x(t)] = 0$, then the variance of $y(t)$ is

$$\sigma_y^2 = E[y(t)^2] = E[\mathbf{a}^H \mathbf{x}(t) \mathbf{x}^H(t) \mathbf{a}] = \mathbf{a}^H \mathbf{R}_{xx} \mathbf{a}, \tag{2.6}$$

where $y(t) = \mathbf{a}^H \mathbf{x}(t)$, H denotes the conjugate transpose operator, and $\mathbf{R}_{xx} = E[\mathbf{x}(t)\mathbf{x}^H(t)]$. In order to design the filter $\mathbf{a}$ to be as selective as possible for the frequency band of interest ω, we can minimize the total power of $y(t)$ subject to the constraint that the filter is undistorted at the frequency ω. Specifically, we can find the filter $\mathbf{a}$ by the following contraint optimization problem:

$$\min \left\{ \mathbf{a}^H \mathbf{R}_{xx} \mathbf{a} \right\} \quad \text{subject to} \quad \mathbf{a}^H \mathbf{e}(\omega) = 1, \tag{2.7}$$

where $\mathbf{e}(\omega) = [1, e^{-j\omega}, ..., e^{-j\omega(p-1)}]^T$. The minimization of (2.7) leads to

$$\mathbf{a} = \frac{\mathbf{R}_{xx}^{-1} \mathbf{e}(\omega)}{\mathbf{e}^H(\omega) \mathbf{R}_{xx}^{-1} \mathbf{e}(\omega)}. \tag{2.8}$$

By substituting (2.8) into (2.6), the power of the resulting signal $y(t)$ after filtering with a bandpass filter $\mathbf{a}$ centered at ω can be obtained by

$$E[y(t)^2] = \frac{1}{\mathbf{e}^H(\omega)\mathbf{R}_{xx}^{-1}\mathbf{e}(\omega)}, \quad 0 \leq \omega \leq \pi. \tag{2.9}$$

According to (2.9), it can be shown in [13] that the spectrum of $x(t)$ can be estimated as

$$S(\omega) \approx \frac{p+1}{\mathbf{e}^H(\omega)\mathbf{R}_{xx}^{-1}\mathbf{e}(\omega)}, \quad 0 \leq \omega \leq \pi, \tag{2.10}$$

where p is the length of filter $\mathbf{a}$.

CHAPTER 3
SUB-OPTIMAL KLT AND ITS REVERSIBLE APPROXIMATION FOR LOSSLESS MULTI-CHANNEL EEG COMPRESSION

3.1 Introduction

Nowadays, EEG has become one of the most useful signals for clinical analysis, i.e. to diagnose the disease and to assess the effectiveness of the treatment via brain functions. However, this analysis process normally takes a very long period of time. Hence, developing the specific type of compression schemes for this signal is an interesting issue. In practice, every sample of the EEG signal is very important and cannot be neglected without the consideration by experts. Therefore, a compression scheme for the EEG signal has to be lossless.

In order to losslessly compress a multi-channel EEG signal, several types of redundancies must be taken into consideration. The temporal redundancy is successfully removed in [14], [15]. Antoniol and Tonella [14] presented a survey on EEG lossless compression algorithms using the predictive coding, transform coding, vector quantization and compared with some well known lossless compression algorithms [16], [17]. Predictive coding are reported to be the most efficient tool that yields an optimal tradeoff between coding performance and computational complexity. In [15], the authors proposed an improved lossless compression scheme on predictive coding by taking the bias cancellation and error modeling into account. Some researchers also employ inter-channel redundancies into consideration but somehow based on the concept of lossy compression.

In [18], Karhunen-Loeve transform (KLT) (also known as principal component analysis (PCA) [19], [20]) is reported to be the best scheme for reducing inter-channel redundancy of the multi-channel ECG compared with the discrete cosine transform (DCT),

discrete wavelet transform (DWT) and lifting wavelet transform (LWT) [21]. The KLT is known for its optimality to diagonalize the autocovariance matrix of an input signal, and to pack the most energy in the fewest transform coefficients, i.e. an input signal is efficiently decorrelated using the KLT. These properties of the KLT are very useful in data compression since less bits will be used to represent the signal [22], [23], [5]. An efficient algorithm employing the KLT to decorrelate inter-channel redundancy of multi-channel signals is also applied to the multi-channel audio coding in [24]. The uses of KLT for reducing the inter-channel redundancy of ECG and audio signals are possible since the number of channels is not likely to be very high. However, for a multi-channel signal such as EEG [4], the number of channels (electrodes) can be as high as a couple hundreds (or a couple thousands in the near future). Calculating the KLT for reducing its inter-channel redundancy is impractical. The problem is even more severe if the KLT is used for lossless compression of EEG (losslessly decorrelated inter-channel redundancy of EEG) since its reversible approximation has to be calculated. Hence, finding a reversible approximation of the KLT with less computational complexity is a challenging problem.

3.1.1 Sub-Optimal KLT and Its Reversible Approximation

Efficient methods for implementing the reversible approximation of the KLT called *reversible KLT* can be calculated by factorizing the KLT matrix into the triangular elementary reversible matrices (TERMs) and single-row elementary reversible matrices (SERMs) [25], [6]. However, the factorization is not unique, and depends on the permutation matrices. Different factorizations yield different dynamic ranges of coefficients, which are of particular importance to lossless compression applications. The large coefficients can significantly impact on the dynamic ranges of the internal nodes in the transform and thus the accuracy of the transform for the same precision. The problem is more severe as the size of the KLT matrix increases.

Considerations on using the reversible KLT for lossless compression include:

- (P1): Since the factorization of the KLT is not unique, finding the best solution in the sense of minimum dynamic range of coefficients is very difficult. An obvious approach is to compare all the possible factorizations. Since the number of solutions is of order $O(N!)$, this, however, is impractical for an N-dimensional signal which N is large.
- (P2): Since the KLT is a statistically dependent transform, its parameters must be transmitted as side information which is of order $O(N^2)$. Hence, the side information should also be minimized.
- (P3): The calculation of the KLT and the implementation of the reversible KLT are highly complex, especially for a high dimensional signal.

In order to solve (P1), the systematic pivoting scheme modified from [2] is proposed by pre-ordering the KLT matrix before applying the suggested pivoting scheme of [2]. The reversible KLT using the proposed pivoting is called *integer KLT* (IntKLT). Furthermore, in order to solve the problems according to (P2) and (P3), an approximated version of the KLT, called *distributed KLT*, is studied in [26] by partially applying KLTs to some selected disjoin groups of divided signals. This study leads to efficient coding performance, however by applying the distributed KLT to a very high dimensional signal, the dimension of divided signals in some groups is still possibly high. This again leads to the difficulties when the distributed KLT is applied to a lossless compression application. Hence, in this thesis, to solve problems (P2) and (P3), an approximation with reversibility preserved for large KLT is proposed. The signals are divided into groups using the 'divide-and-conquer' philosophy. For each group, the signals are decorrelated by their fixed smaller-size KLT (referred to as *marginal KLT* in [26]). Large-magnitude transform coefficients are then grouped and further decorrelated using their marginal KLT. Small-magnitude transform coefficients are passed through the next stage. By

repeating this self-similar structure, a novel transform called sub-optimal KLT is constructed. Each of the marginal KLTs of the sub-optimal KLT can be further realized by a lifting factorization, rendering a reversible transform with small rational lifting coefficients. For convenience, this reversible approximation of the sub-optimal KLT is called *integer sub-optimal KLT* (IntSKLT). As mentioned, in order to obtain small lifting coefficients, appropriate permutation matrices have to be well selected. Finding the right permutation matrices for factorizing each smaller-sized marginal KLT instead of full-sized N-point KLT is more practical. As a result, using the IntSKLT can significantly reduce the computational complexity while keeping a comparable coding gain. In practice, the IntSKLT can be used to reduce the redundancy of a high dimensional signal in a lossless coder.

3.1.2 Lossless Multi-Channel EEG Compression

The existing literatures are still lacking of a systematic approach for losslessly compressing multi-channel EEG, i.e. losslessly reducing both the inter-channel and temporal redundancy. In this thesis, to solve this problem and to evaluate the usefulness of our proposed IntSKLT, we also propose a lossless compression algorithm for the multi-channel EEG. In order to shift the mean of each channel back to zero, we employ a simple backward difference along the temporal domain of each channel. Since the number of channels of EEG can sometimes be very high, to optimally reduce the inter-channel correlation, the IntSKLT is employed. The temporal redundancy of each channel is then reduced by employing the predictive coding or the integer-to-integer mapping approximation of the DCT type IV (stereo IntDCT-IV) [27]. The coding results show that the proposed lossless coder using the IntSKLT outperforms the benchmark lossless coders by more than 20% in compression ratio. The publication related to this chapter can be found in [28], [29], and [30].

3.2 Sub-Optimal KLT and Its Reversible Approximation

This section presents a novel sub-optimal KLT. Its performance when the input is the 1-st order Markov process is studied in Section 3.2.1 and 3.2.2. In order to use the sub-optimal KLT for lossless compression, its reversible structure called IntSKLT is also presented in Section 3.2.3. The performance of the IntSKLT is evaluated in Section 3.2.4.

3.2.1 Sub-Optimal KLT

In order to solve the issue caused by applying the KLT to a high dimensional signal (say N dimensions), an N-dimensional signal (input of the KLT) is equally divided into two groups of $N/2$ points (assume that N is even) along the arbitrary scan order (Fig. 3.9). Each group is decorrelated by its local marginal KLT, and the outputs are sorted according to the variances from large to small. To further reduce the dependency, the largest $N/4$ points from the two groups are combined, and decorrelated by their $N/2$-point KLT. Under the assumption that, the input signals are highly decorrelated after the first stage, the remaining $N/2$ points of KLT coefficients, which are usually very small, are exported directly to the outputs. Fig. 3.1 shows a block diagram for the case of $N = 64$. A self-similar structure is used to further reduce the complexity where the same structure is repeated for each of the $N/2$ KLT. The details of each 32-point sub-optimal KLT in Fig. 3.1 is illustrated in Fig. 3.2, where the iterations of dividing the signals stop at *8-point KLT*, which is a feasible size to optimize for the best parameters. It should be noted that optimality of the proposed approximation of the KLT also depends on the statistics of the signals and how they are grouped at the first stage.

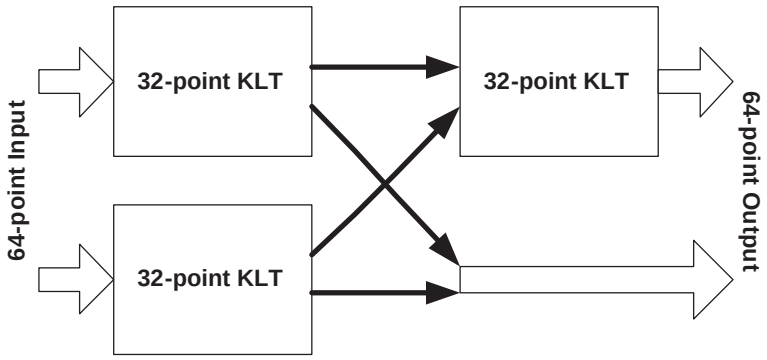

Figure 3.1. Structure of the 64-point sub-optimal KLT.

3.2.2 Markov Process

In this section, let us consider a special case of Markov process. Assume that the correlation between the i-th and j-th points of a signal is given by:

$$[\mathbf{R}_{xx}]_{ij} = E\{x_i x_j\} = \rho^{|i-j|}, \quad i, j = 0,\ 1, \cdots,\ N-1.$$

It can be shown that the maximum achievable coding gain [31] for this case is

$$G_N = \frac{\frac{1}{N} tr(\mathbf{R}_{xx})}{det(\mathbf{R}_{xx})^{1/N}} = (1-\rho^2)^{-(1-\frac{1}{N})},$$

where tr denotes the trace operation.

3.2.2.1 4-point Case

Consider a simple case of $N = 4$ where the signals are decorrelated using the structure in Fig. 3.3. It can be shown that the marginal KLTs $\mathbf{T}_1$ and $\mathbf{T}_2$ are

$$\mathbf{T}_i = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}, \tag{3.1}$$

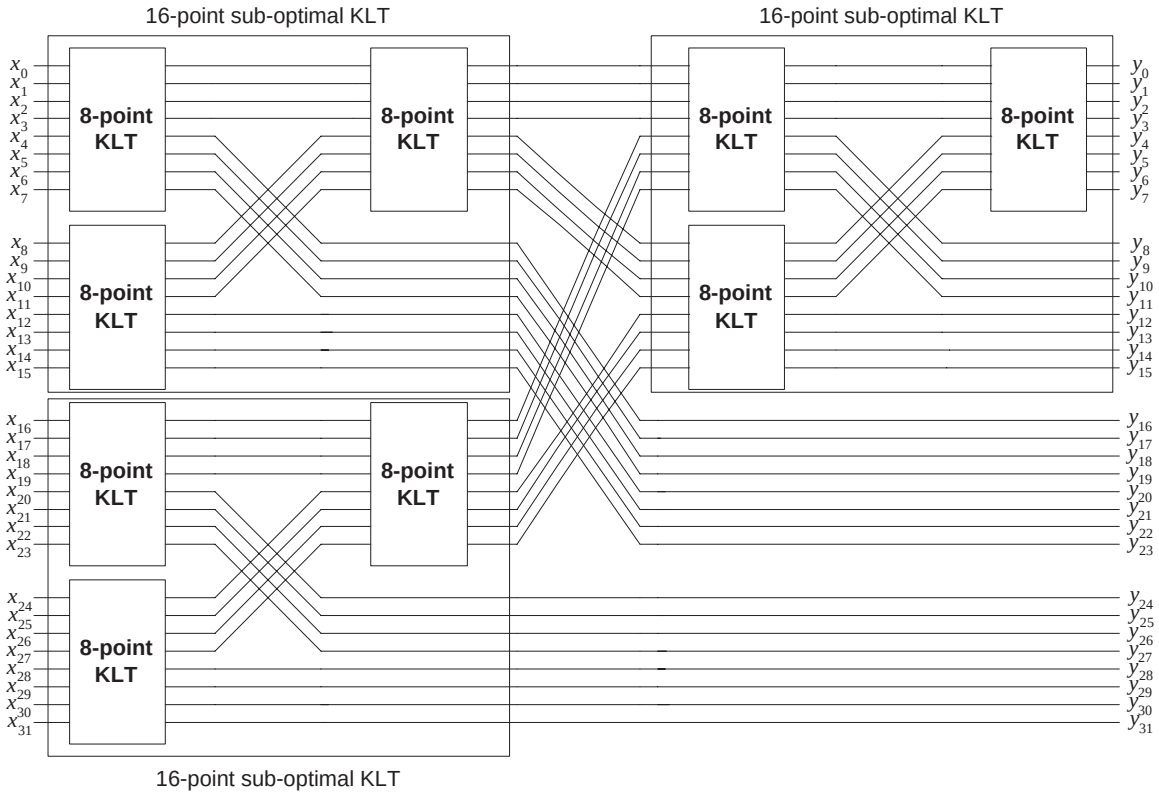

Figure 3.2. Structure of the 32-point sub-optimal KLT which is composed of the 16-point sub-optimal KLTs and the 8-point KLTs.

and the corresponding eigenvalues are $1 \pm \rho$. The coding gain after the first stage (z_i) is $(1-\rho^2)^{-\frac{1}{2}}$. Applying a 2-point KLT to the larger components z_0 and z_2 results in a sub-optimal coding gain of

$$\hat{G}_4 = (1-\rho^2)^{-\frac{1}{2}} \left[1 - \frac{1}{4}\rho^2(1+\rho)^2\right]^{-\frac{1}{4}}.$$

It can be shown that $0.9036 \approx \left(\frac{2}{3}\right)^{1/4} \leq \frac{\hat{G}_4}{G_4} \leq 1$ for $0 \leq \rho \leq 1$.

3.2.2.2 2^p-point Case

In the extension to the case of $N = 2^p$, assume that the three $N/2$-point KLTs used in the approximation in Fig. 3.1 are obtained by the eigenvectors of the correlation matrices of the corresponding inputs, i.e. the three KLTs are the marginal KLTs. Fig. 3.4

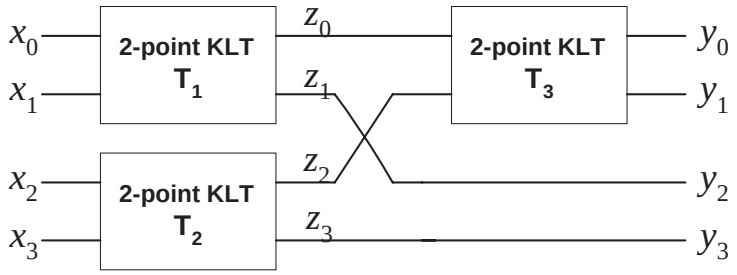

Figure 3.3. Structure of the 4-point sub-optimal KLT.

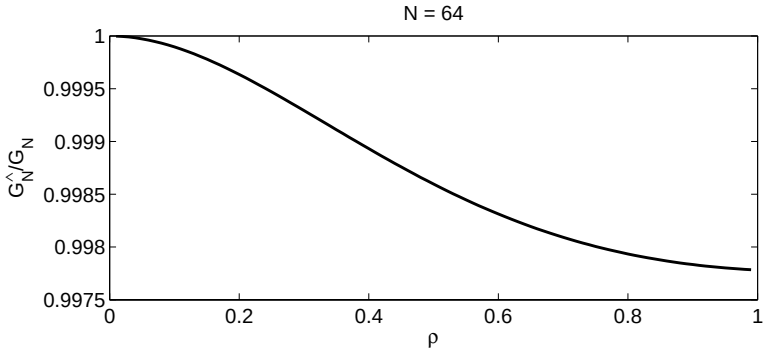

Figure 3.4. $\hat{G}_N/G_N$ for the case of $N = 64$.

plots the ratio $\frac{\hat{G}_N}{G_N}$ as a function of ρ for the case of $N = 64$. It is clear that when these sub-KLTs are optimal, the degradation in coding gain is very insignificant (less than 0.3%) compared to the original KLT.

In order to apply the proposed sub-optimal KLT to the compression scheme, two points should be noted. First, in the proposed recursive structure, the $N/2$-point KLTs are further approximated, and thus the difference in coding gains accumulates. Second, and perhaps more importantly, EEG signals may not be a Markov process. In fact, the correlation structure can be in 2-D (e.g. for a multi-channel EEG, one channel may be correlated to all of its surrounding neighboring channels). Hence, the choice in clustering the inputs into groups also has an impact on the coding performance. The preliminary study on this issue is demonstrated in Section 3.2.4.

3.2.3 Integer Sub-Optimal KLT (IntSKLT)

This section illustrates the structure of the IntSKLT. Since the IntSKLT is the transform that composes of many blocks of IntKLTs, we will first describe the details of the IntKLT followed by the IntSKLT.

3.2.3.1 IntKLT

According to Section 2.3, any $N \times N$ non-singular matrix $\mathbf{A}$ with determinant of ± 1 can be factorized as in (2.2). Thus, $\mathbf{A}$ can be approximated with reversibility preserved by simply quantizing the off-diagonal of $\mathbf{S}_i$. As mentioned in Section 3.1, the factorization in (2.2) is not unique and depends mainly on the permutation matrices. To demonstrate this, let us consider an example.

Let $\mathbf{A}$ be an orthogonal matrix of size 3×3:

$$\mathbf{A} = \begin{bmatrix} 0.6 & -0.64 & 0.48 \\ -0.8 & -0.48 & 0.36 \\ 0 & 0.6 & 0.8 \end{bmatrix}.$$

Two lifting factorizations of $\mathbf{A}$ can be given by:

$$\begin{aligned} \mathbf{A} = & \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0.33 & -0.5 & -1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0.27 & 1 & -0.8 \\ 0 & 0 & 1 \end{bmatrix} \\ & \begin{bmatrix} 1 & -0.3 & 0.6 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -0.33 & 0.5 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \text{or} \end{aligned} \tag{3.2}$$

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 3.75 & -3.42 & -1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 1.2 & 1 & -0.48 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & -3.33 & 0.8 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1.25 & 4.92 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \quad (3.3)$$

It is clear that both (3.2) and (3.3) are equal, and can be approximated with reversibility preserved. The magnitudes of the coefficients in (3.2) are all less than or equal to one whereas those in (3.3) are as high as 4.92.

In general, the factorization in (2.2) leads to high dynamic range of the matrix elements which has a direct impact on lossless coding performance. Intuitively speaking, if the elements in (2.2) are greater than 1, when the rounding operators are applied to each internal node (Fig. 3.6), the rounding errors are amplified. Consequently, the lossless coding performance is degraded. The traditional procedure to solve this problem is to find all possible factorizations so that the rounding error is minimized. However, this scheme is impractical due the large number of solutions which is of order $O(N!)$. This problem can be solved by carefully choosing the permutation matrices using the efficient pivoting scheme.

- *Proposed Pivoting Scheme*: In [2], a pivoting scheme for reducing the dynamic range of the factorized matrix coefficients is proposed. A condition on how to select the permutation matrix $\mathbf{P}_L$ in (2.2) is simply added in the factorization process. The permutation matrix $\mathbf{P}_L$ is selected such that only the coefficients in $\mathbf{S}_0^T$ are minimized. It should be noted that the possible factorization solutions is now down to order $O(N)$. This pivoting scheme results in a similar factorization to (2.2) except that $\mathbf{S}_0$ is now a lower-triangular matrix with diagonal elements of

± 1. Although, using the factorization in (2.2) together with the pivoting scheme in [2] seems to work well, there still be an issue in lacking of the degree of freedom on column interchange which may result in unexpected large dynamic range of the coefficients in some cases.

Without losing any information, we simply propose the solution to the above problem by factorizing the column-interchanged version of $\mathbf{A}$ instead of $\mathbf{A}$ itself. In general, another permutation matrix $\mathbf{P}_R$ is incorporated to (2.2) as

$$\mathbf{A}_{N\times N} = \mathbf{P}_L\mathbf{S}_N\mathbf{S}_{N-1}\cdots\mathbf{S}_1\mathbf{S}_0\mathbf{P}_R. \tag{3.4}$$

Heuristically, the right permutation matrix (column pre-ordered matrix) $\mathbf{P}_R$ is suggested to be a rotated version of the identity matrix, therefore we have only N choices of $\mathbf{P}_R$. Generally, the pivoting scheme in [2] is only a special case of the proposed pivoting scheme. In comparison, the proposed pivoting scheme introduces slightly more possible factorization solutions of order $O(N^2)$. However, these are much fewer than the solutions of the traditional procedure which is of order $O(N!)$. Furthermore, it is evident in Fig. 3.5 that by using the proposed pivoting scheme to parameterize the IntKLT, the rounding error is approximately reduced by almost 5%.

- *Structure of the N-point IntKLT*: Since the KLT is an orthonormal transform, its determinant is ± 1. To obtain a reversible approximation of the factorization in (3.4), a rounding operator is applied at each lifting stage, resulting in the IntKLT. Fig. 3.6 illustrates the structure of the proposed IntKLT for the case of $N = 4$ (4-point IntKLT). As one can see, there are a total of $\frac{3}{2}N(N-1)$ coefficients and only $2N-1$ rounding operations. Although the proposed factorization can efficiently approximate the KLT, factorizing a large KLT matrix is still impractical. .

3.2.3.2 IntSKLT

Once the sub-optimal KLT is obtained (as in Section 3.2.1), it is very simple to implement its reversible realization. Each marginal KLT matrix in the structure of the sub-optimal KLT is replaced with its corresponding IntKLT rendering a reversible approximation of the sub-optimal KLT called IntSKLT. For example, the reversible approximation of the 32-point sub-optimal KLT can be constructed by replacing each 8-point KLT in Fig. 3.2 with the 8-point IntKLT.

The computational complexity, i.e. the number of lifting coefficients, of the N-point IntSKLT remains only approximately $3.1N^{1.585}$ ($\frac{3}{2}8(8-1) \times 3^{\log_2(N/8)} = 84 \times 3^{\log_2(N/8)} = \frac{28}{9}3^{\log_2 N} = \frac{28}{9}N^{\log_2 3} \approx 3.1N^{1.585}$) compared with $\frac{3}{2}N(N-1)$ used in the N-point IntKLT. It is clear that the difference between the number of coefficients from both cases gets larger especially when N is large. In general, using the IntSKLT can reduce the computational complexity from using the IntKLT by more than 60%. In addition, according to the fact that the IntSKLT is not an ideally optimal transform, theoretical optimal coding property, i.e. the optimal energy compaction [5], will not be as efficient as the IntKLT. However, since the IntSKLT uses less amount of transform coefficients, i.e. less amount of side information, its lossless coding performance will be just slightly degraded from the IntKLT. This will be further clarified by the simulation results in Section 3.3.

3.2.4 Transform Evaluation

Fig. 3.7 compares the (sorted) output variances resulting from the IntKLT, IntSKLT and IntDCT [32] when being applied to decorrelate the inter-channel redundancies of the multi-channel EEG signal in Fig. 3.8(a). Since the IntDCT is signal independent, there are more large residual outputs compared with the IntKLT, while the output variances of the IntSKLT are very similar to those obtained from the IntKLT.

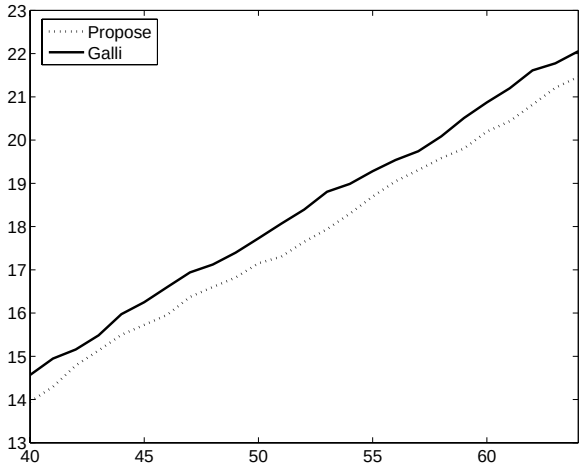

Figure 3.5. Cumulative mean absolute error of the proposed pivoting and the pivoting in [2] for the 64-point IntKLT using 8,192 samples of EEG signals, where x-axis represents the number of cumulated points and y-axis represents the cumulative mean absolute values.

In this experiment, the order of the input to each marginal KLT (of the IntSKLT) is also taken into consideration. We rearrange the multi-channel EEG signal according to two types of channel scans (channel ordering), i.e. spiral scan (Fig. 3.9(a)) and clustering scan (Fig. 3.9(b)). Figs. 3.7(a) and (b) show that by using the clustering scan, the output variances obtained by the IntSKLT are closer to those of the IntKLT than those using the spiral scan. Both scanning schemes are proposed based on the assumption that, the neighboring channels are similar to each other. The spiral scan is constructed by scanning the channels from the center channel (CZ) in a circular way to the outer channel (VEOU). The clustering scan is constructed by grouping the neighboring channels together. It should be noted that the patterns of both scans are empirically constructed in order to group the signals that have the similar structures together. The optimal choice of scan needs further investigation.

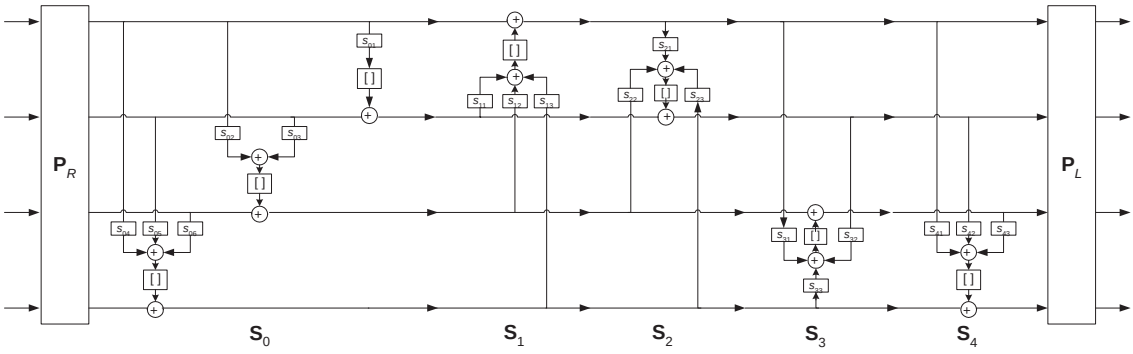

Figure 3.6. Reversible structure of the factorization using (3.4) for the case of 4×4 KLT matrix (4-point IntKLT).

3.3 Application of the IntSKLT in Lossless Multi-channel EEG Compression

In this section, the lossless coder for a multi-channel EEG is presented. Fig. 3.10 shows a block diagram of the proposed coder consisting of four main stages: preprocessing, inter-channel decorrelation, temporal decorrelation and entropy coding.

3.3.1 Preprocessing

According to a data acquisition process, strengths of the potentials at different electrode positions on the scalp can be different. Figs. 3.8(a)-(d) show various EEG waveforms measured from the neighboring locations. It is observed that, while sharing some temporal similarities, they have different DC-bias. This bias complicates and degrades the inter-channel decorrelation performance. Removing the DC-bias can simplify the KLT and hence the IntKLT and IntSKLT used in the inter-channel decorrelation stage. Thus, a simple backward difference is applied along the temporal domain of each channel as a preprocessing:

$$y_i[n] = x_i[n] - x_i[n-1],$$

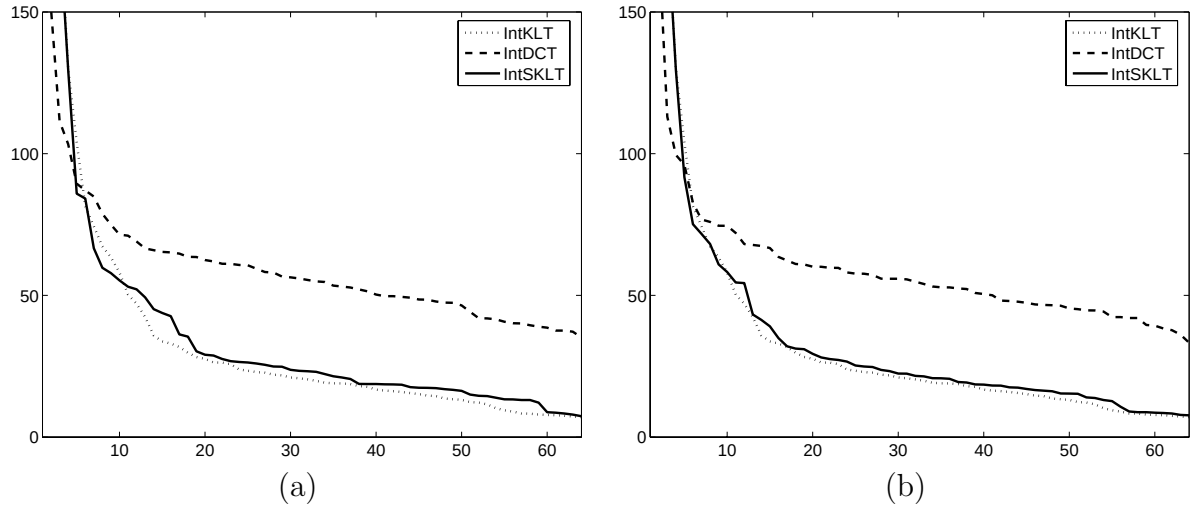

Figure 3.7. Performance analysis of the 64-point IntKLT, IntDCT and IntSKLT using: (a) spiral and (b) clustering scanned versions of data 1, where x-axis represents the transform coefficients and y-axis represents their (sorted) variances.

where i indicates the channel index ranging from 1 to N, N is the number of channels. In summary, this process results in zero-mean signals $y_i[n]$, and thus improves the coding performance.

3.3.2 Spatial (Inter-Channel) Decorrelation

In this process, we aim to reduce the inter-channel redundancy of a multi-channel EEG signal. In general, the inter-channel correlation of a multi-channel EEG signal is location-dependent. Therefore, fixed-basis transforms such as DCT, DFT and LWT may fail to efficiently exploit the inter-channel relationship. To alleviate this problem, a data-dependent type of transforms is preferred. Hence, in our proposed coder, the KLT which is optimal and data dependent is used to reduce the inter-channel redundancy of the multi-channel EEG signal. In particular, let $\mathbf{T}$ be the KLT matrix obtained from the multi-channel signal then

$$\mathbf{y}[n] = \mathbf{T}\mathbf{x}[n],$$

Table 3.1. Compression ratios (CR) when different inter-channel decorrelation methods are applied (NT denotes No Temporal decorrelator).

Coder	data1	data2	data3	data4
IntKLT+NT	2.37	2.38	2.39	2.24
IntSKLT+NT	2.35	2.37	2.37	2.20
IntDCT+NT	2.12	2.12	2.10	1.97
LWT+NT	2.16	2.17	2.17	2.02
AR(1)+NT	2.22	2.26	2.24	2.12
Prediction transform+NT	2.22	2.24	2.23	2.07
IntKLT+AR(6)	2.94	2.97	2.97	2.74
IntSKLT+AR(6)	2.93	2.95	2.94	2.71
IntDCT+AR(6)	2.67	2.71	2.69	2.49
LWT+AR(6)	2.50	2.51	2.52	2.32
AR(1)+AR(6)	2.73	2.75	2.75	2.58
Prediction transform+AR(6)	2.77	2.78	2.78	2.58

where $\mathbf{x}[n] = [x_1[n], \ldots, x_N[n]]^T$ denotes the vector of the n-th samples of the N-channel data. $\mathbf{y}[n] = [y_1[n], \ldots, y_N[n]]^T$ denotes the corresponding KLT coefficients of $\mathbf{x}[n]$, and the KLT matrix $\mathbf{T}$ is calculated by $\frac{1}{N}\sum_n \mathbf{x}[n]\mathbf{x}[n]^T$. To losslessly decorrelate the spatial correlation, the KLT can be approximated using the IntKLT or IntSKLT.

In general, when the IntKLT or IntSKLT are used as the channel decorrelator, the energy will be compacted in the very first channels and very small in the other channels. This results in an efficient coding performance [5]. For convenience, the resulting channel after applying the IntKLT or IntSKLT will be called eigenchannel.

3.3.3 Temporal Decorrelation

After spatial decorrelation, temporal redundancy is exploited separately in each eigenchannel. Considerations for the choice of transform in this process include that:

Table 3.2. Performance analysis (in term of CR) of using the IntKLT and the IntSKLT with different scanning types.

64-channel EEG signals	data1	data2	data3	data4
NC+AR(6)	2.63	2.63	2.62	2.39
IntKLT+AR(6)	2.94	2.97	2.97	2.74
IntSKLT(clustering scan)+AR(6)	2.93	2.95	2.95	2.72
IntSKLT(spiral scan)+AR(6)	2.91	2.93	2.92	2.67
IntSKLT(random scan)+AR(6)	2.86	2.92	2.92	2.64

1. it can efficiently reduce the temporal redundancy,
2. it is a reversible process due to the lossless issue, and
3. it can be computed at low cost.

In this thesis, two classical temporal decorrelators are employed. By assuming the EEG signal as time series, one of the classical predictive coding, i.e. the 6-order autoregressive model (AR(6)), is used [15]. Actually, the AR(6) can be further improved by using together with the context-based bias cancellation (a context based error correction) [15]. However, to fairly compare with some other existing lossless algorithms, we drop this error correction part in our lossless coder. By assuming that the signal is quasi-nonstationary, the non-overlapping sliding window of the stereo IntDCT-IV proposed in [27] as depicted in Fig. 3.11 is employed as a temporal decorrelator of each eigenchannel (In our experiment, the window is slided by 1,024 samples). The key idea of the stereo IntDCT-IV is that we can decorrelate two channels at the same time while obtaining the reversible structure via lifting scheme. Furthermore, by employing two channels at the same time, the rounding operators as well as the rounding error are reduced. The reader is referred to [27] for the proof of the structure in Fig. 3.11.

Table 3.3. Compression ratios obtained from using proposed pivoting (denoted as proposed) and pivoting scheme in [2] (denoted as Galli) for factorizing the KLT matrices.

Coder	data1	data2	data3	data4
IntKLT+stereo IntDCT-IV (Proposed)	2.82	2.84	2.83	2.63
IntKLT+stereo IntDCT-IV (Galli)	2.81	2.83	2.82	2.63
IntKLT+AR(6) (Proposed)	2.94	2.97	2.97	2.74
IntKLT+AR(6) (Galli)	2.94	2.96	2.96	2.74
IntSKLT+stereo IntDCT-IV (Proposed)	2.83	2.86	2.85	2.66
IntSKLT+stereo IntDCT-IV (Galli)	2.83	2.85	2.84	2.66
IntSKLT+AR(6) (Proposed)	2.93	2.95	2.94	2.71
IntSKLT+AR(6) (Galli)	2.87	2.94	2.94	2.71

3.3.4 Entropy Coding

Huffman coding is used for two main reasons:

1. to further losslessly reduce the statistical redundancy, and
2. to fairly compare the coding results with other lossless algorithms.

It is noted that more sophisticated entropy coding, e.g. the activity-based conditional coding [15], can also be applied in order to achieve a better coding performance.

Coding Results

In this simulation, the proposed lossless coder is applied to 8 seconds of 4 datasets of 64-channel EEG in Fig. 3.8, e.g. left eyeblink, left-right eye blink, eye rotating and speaking (called data 1-4, respectively). All data is sampled at 1.024 kHz and digitized to 16 bits, i.e. each channel has 8,192 samples. KLT matrices and AR coefficients obtained in these experiments are trained from the first 2,048 samples of each channels of each dataset.

Table 3.4. Compression ratios obtained from applying various types of lossless coders to data 1-4 (NC denotes No Channel decorrelator and NT denotes No Temporal decorrelator).

Coder	data1	data2	data3	data4
IntKLT+stereo IntDCT-IV	2.82	2.84	2.83	2.63
IntKLT+AR(6)	2.94	2.97	2.97	2.74
IntSKLT+stereo IntDCT-IV	2.83	2.86	2.85	2.66
IntSKLT+AR(6)	2.93	2.95	2.94	2.71
NC+stereo IntDCT-IV	2.53	2.54	2.52	2.31
NC+AR(6)	2.63	2.63	2.62	2.39
NC+NT	2.07	2.17	2.04	1.89
Huffman	1.34	1.28	1.39	1.27
Shorten	2.24	2.03	1.93	1.70
JPEG2000	2.07	2.07	2.06	2.05
GZIP	1.44	1.43	1.38	1.29

Table 3.1 illustrates the efficiency of various inter-channel decorrelation schemes, e.g. the IntKLT, IntSKLT, IntDCT [32], LWT [21], AR(1) (backward difference along channels), and prediction transform. The prediction transform used for the comparison in this experiment corresponds to a reversible triangular matrix shown in Fig. 3.12, where the coefficients s_k are selected to minimize the mean square error of the prediction. In particular, the prediction transform is simply a linear prediction along the channels which its structure is a special case of the IntKLT. The backward difference is applied (along the temporal domain) to each channel of EEG prior to the channel decorrelation as the preprocessing stage (Fig. 3.10). By using these schemes as the channel decorrelators without any temporal decorrelator, maximum compression ratios are obtained by using the IntKLT and IntSKLT while using the AR(1), prediction transform, LWT and IntDCT yield the degradation in compression ratios. Furthermore, by using the IntKLT,

IntSKLT, IntDCT, LWT, AR(1) and prediction transform as the channel decorrelators and using the AR(6) as the temporal decorrelator, Table 3.1 also illustrates that maximum compression ratios are still obtained by using the IntKLT and IntSKLT while using the prediction transform, AR(1), IntDCT and LWT yield the degradation in compression ratios.

Table 3.2 compares the performances of the IntKLT and IntSKLT with different scanning types. Therefore, in all cases, AR(6) is used as the temporal decorrelator. When there is no inter-channel decorrelator, i.e. the spatial decorrelation block in Fig. 3.10 is removed, the coder yields the worst compression ratio. The maximum performances are achieved by using the IntKLT which is slightly higher than using the IntSKLT. By applying the clustering, spiral and random scans to the input signals of the 64-point IntSKLT, the compression ratios are respectively degraded. This shows that the order of EEG channels has some impacts on the coding performance of the IntSKLT.

Table 3.3 illustrates the merit of the proposed pivoting scheme (denoted as Proposed) over the pivoting scheme in [2] (denoted as Galli). By using the proposed coder with different inter-channel decorrelators (e.g. IntKLT and IntSKLT) and temporal decorrelators (e.g. stereo IntDCT-IV and AR(6)), the proposed pivoting scheme leads to the improvement in coding performance especially in data1.

Finally, Table 3.4 compares the coding performances of our proposed lossless coders (using the IntKLT and IntSKLT as the inter-channel decorrelators, and using the stereo IntDCT-IV and AR(6) as the temporal decorrelators) with the benchmark lossless coders such as

- Structure in Fig. 3.10 except that the inter-channel decorrelator is discarded and stereo IntDCT-IV is used as the temporal decorrelator (NC+stereo IntDCT-IV, NC denotes No Channel decorrelators),

- Structure in Fig. 3.10 except that the inter-channel decorrelator is discarded and AR(6) is used as the temporal decorrelator (NC+AR(6)),
- Structure in Fig. 3.10 except that both inter-channel and temporal decorrelators are discarded (NC+NT, NT denotes No Temporal decorrelators), i.e. using only preprocessing and Huffman coding,
- Huffman coding,
- GZIP [17],
- Optimal linear prediction based lossless coding (Shorten) [33],
- Lossless JPEG2000 [34]

For all data, the proposed coder yields the highest compression ratios. Even though, using the IntSKLT would lead to slightly degradation in coding performance compared with using the IntKLT, we can achieve more than 80% reduction in *computation time* when performed on the laptop computer with a 1.6 GHz Intel Pentium M processor. Although the classical lossless coding scheme such as GZIP yields less time-complexity than the proposed coder, since EEGs do not have much frequent reoccurrence data, using GZIP would not be a good candidate coder [15]. Also, by using Shorten, Lossless JPEG2000, only AR(6) and only stereo IntDCT-IV, inter-channel redundancies have not been taken into consideration. Using the stereo IntDCT-IV as the temporal decorrelator yields slightly degradation in coding performance compared with the AR(6). However, since the stereo IntDCT-IV is signal independent, it can be useful in some cases where the EEG signals are very small and thus the calculation of the AR coefficients is ill-posed.

In this experiment, KLT matrices of the size 64×64 for 64-channel EEG have to be transmitted as the side information. Since the KLT matrices offline calculated from the training set can be used as the fixed transforms for the online stage, less than 3% of the overall transmitted bits are used as the side information for the IntKLT and less than 1.5% for the IntSKLT.

3.4 Summary

In this chapter, we have presented a novel sub-optimal KLT. This transform can eliminate the complexity issue when applying the KLT to a high dimensional data. Based on this sub-optimal KLT which composes of many sub-KLTs, we show that the reversible approximation, called IntSKLT, can be implemented easily since we only have to deal with the factorization of smaller-sized matrices. A new pivoting scheme is also presented to further improve the performance of the reversible structure. The results show that the proposed pivoting scheme can efficiently improve the coding performances of the IntKLT and IntSKLT. Moreover, a lossless coder designed for multi-channel EEG which exploits the inter-channel and temporal correlation is presented. Since the IntSKLT yields coding gain close to the original KLT while requiring less side information for perfect reconstruction, this lossless coder is very useful especially when the number of channels is high. The improvements in coding performances of the proposed coder compared with the benchmark lossless coders are also illustrated.

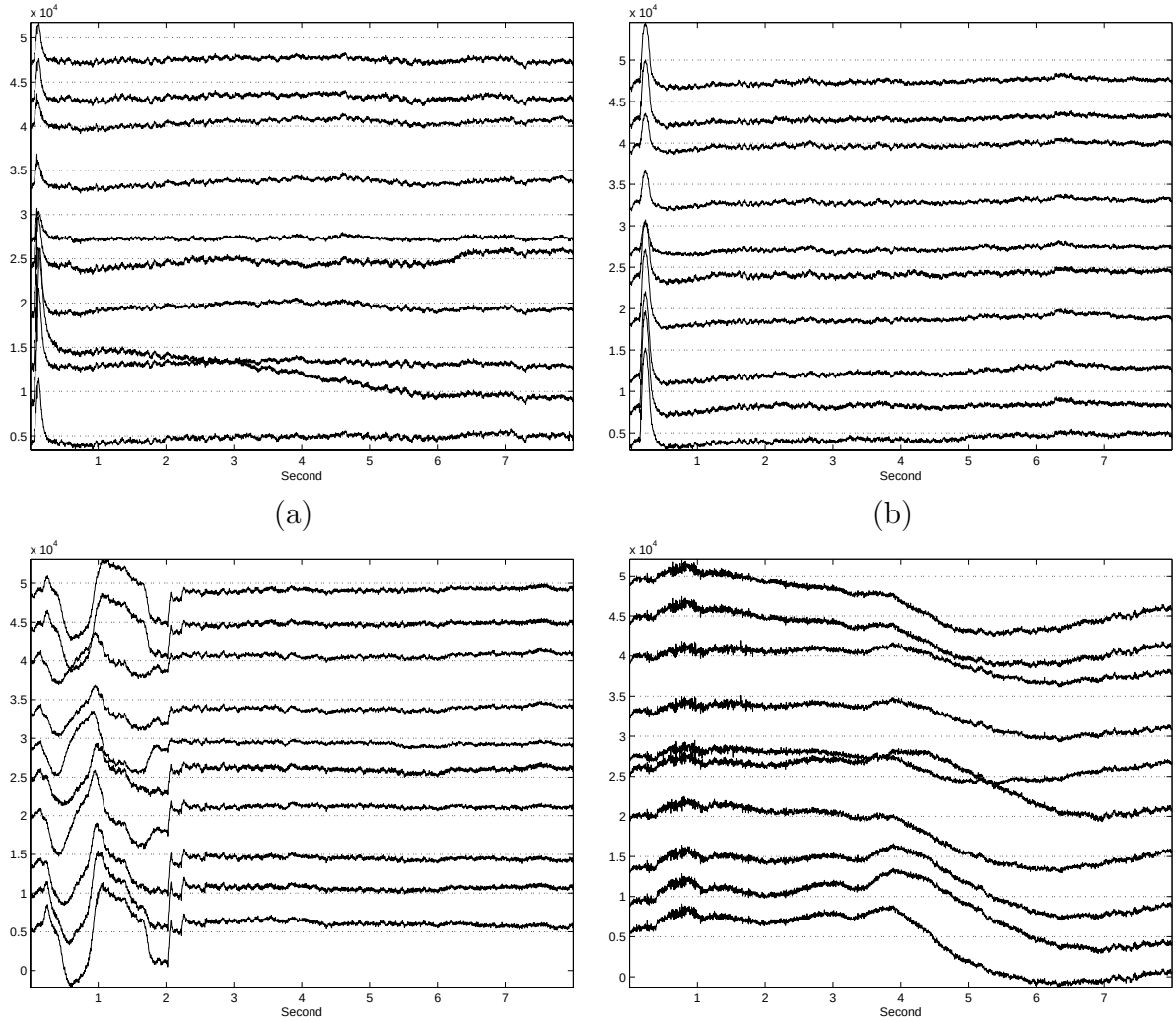

Figure 3.8. The first 10 channels of 4 datasets according to the clustering scan (from the bottom to the top) in Fig. 3.9(b): (a) Left eyeblink (data1), (b) Left-right eyeblink (data2), (c) Eye rotating (data3) and (d) speaking (data4). x-axis represents the time (in second) and y-axis represents the size of each EEG data. The dot lines represent the zero-mean line for each channel of EEGs.

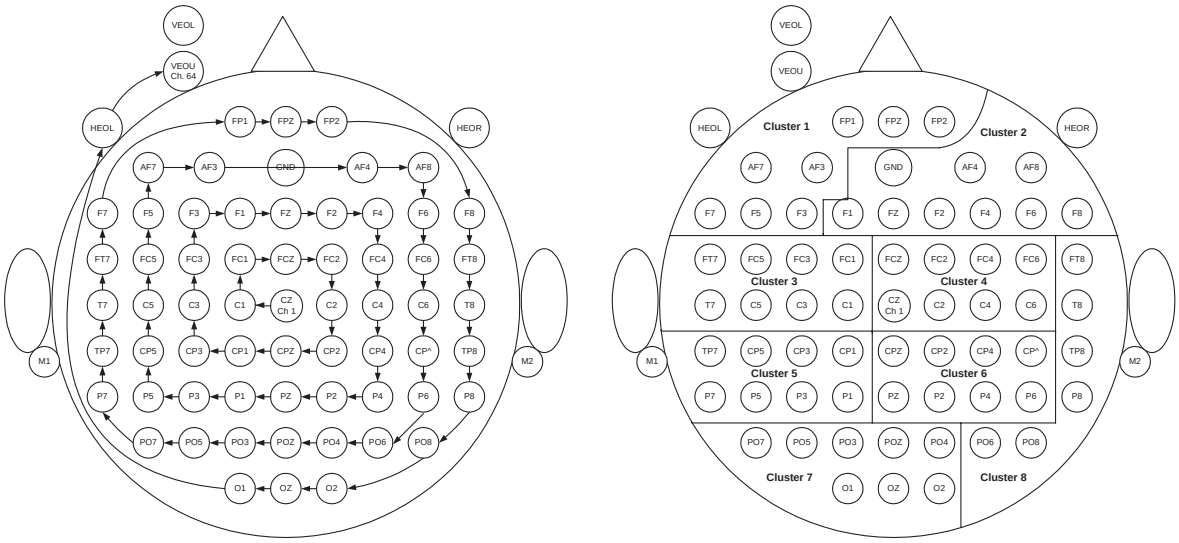

Figure 3.9. Pattern of scanning schemes: (a) spiral scan and (b) clustering scan (Channel HEOL and VEOU are included in cluster 8).

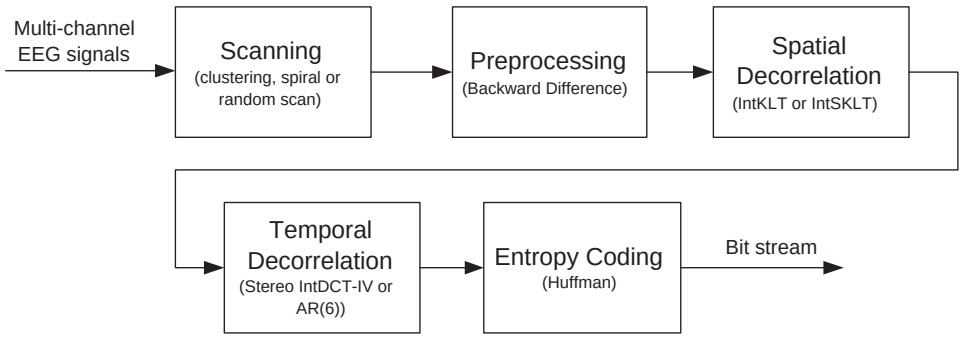

Figure 3.10. Block diagram of the proposed lossless coder.

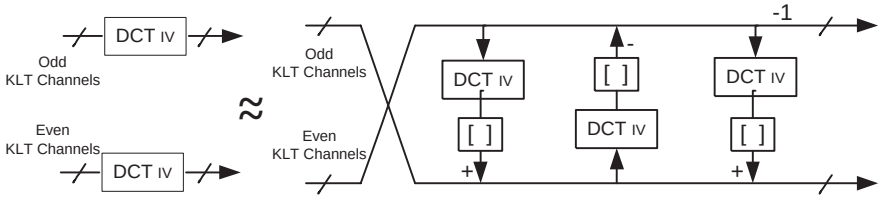

Figure 3.11. Ladder structure of the stereo IntDCT-IV.

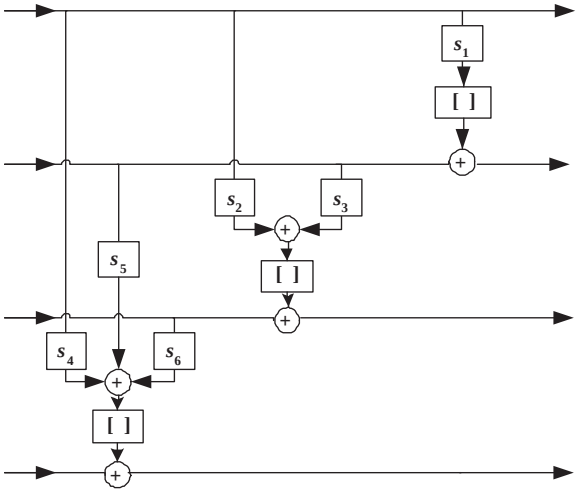

Figure 3.12. Reversible structure of the prediction transform used for inter-channel decorrelation of a 4-channel EEG signal.

CHAPTER 4
REDUCED COMPLEXITY SPACE-TIME-FREQUENCY MODELS FOR MULTI-CHANNEL EEG

4.1 Introduction

In order to efficiently exploite a multi-channel EEG in real world applications, e.g. brain computer interface (BCI) [35], classification, analysis, and prediction, we have to fully understand the pattern of this multi-channel signal. To achieve this, all information should be incorporated to form an efficient model. Therefore, finding a right model to extract the features of the multi-channel EEG with less time consuming becomes one of the challenging problems in neuroscience.

EEG is first modeled by its frequency statistics in [36]. The model is further improved by using time-frequency representation of a single channel EEG, [37], [38], [8] which is known as a nonstationary signal. Usually, EEG signals are recorded at multiple locations, yielding information about which part of the brain is functioning. This spatial knowledge is efficiently exploited using principal component analysis (PCA) in [39], [40]. However, by using PCA nonuniqueness occurs due to arbitrary choice of rotational axes [10], which leads to the robustness problem of the model. Recently, independent component analysis (ICA) is applied to eliminate this nonuniqueness problem by imposing the statistical independent constraint which is even stronger than orthogonality of PCA, [41], [42]. In conventional PCA and ICA, no frequency knowledge is exploited even though it can be separately employed later. All space, time, and frequency domains are employed in [43] by analyzing the region of time-frequency plane. Another interesting work on topographic-time-frequency decomposition is proposed in [44] by imposing the minimum norm and maximal smoothness to the time and frequency signatures, respec-

tively, for uniqueness of the model. Recently, Miwakeichi *et al* [10] found that by using PARAFAC, [7], [45] these uniqueness constraints are unnecessary. Therefore, they propose a novel model that applies space-time-frequency representation of a multi-channel EEG to a 3-way PARAFAC to obtain the space, time and frequency signatures (features), called space-time-frequency model (STF model). Although, all domains are exploited in these models, they suffer from the high computational complexity when measured in a long period of time or with high number of electrodes.

In this chapter, we present three methods to reduce the computational complexity of the STF model for a multi-channel EEG. The first method aims to estimate the STF model using the space-time-frequency-time/segment model (STF-TS model) by subdividing the time domain into a number of segments resulting in a 4-D array signal. The 4-way PARAFAC is then applied for the analysis of the 4-D array signal. This approach is appropriate when the signals are recorded for a long period of time. The second method aims to estimate the STF model using the space-time-frequency-space/segment model (STF-SS model), which is suitable when the number of channels (dimension of space domain) is high. By partitioning the channels into sub-groups, a 4-D array signal is constructed, and the 4-way PARAFAC is then applied for the analysis. However, if the dimensions of both time and space domains are high, the computation of these models can be further reduced. Therefore, we extend the idea by simultaneously partitioning the multi-channel EEG in both space and time domains called space-time-frequency-space/segment-time/segment model (STF-SS-TS model). These three reduced complexity models are further shown to be useful in many EEG analysis problems, e.g. artifact removal and classification of long-term and high number of channels EEG signals. The content of this chapter is published in [46], [47], [48], and [49].

4.2 Reduced Complexity Space-Time-Frequency (STF) Models

By segmenting the selected domains (space, time or both) the STF model with the additional domains called space/segment and time/segment can be obtained. In this section, three modeling schemes based on this '*divide and conquer*' philosophy are introduced.

4.2.1 Space-Time-Frequency-Time/Segment Model (STF-TS model)

For a long-term EEG signal, the calculations of both the time-frequency transform and PARAFAC are very complex. Therefore, we aim to reduce this computational complexity by dividing the time domain into segments. After that, the time-frequency transform is applied individually to each segment forming a 4-D array signal, $y(n_1, s_t, f_1, t_1)$ (denoted in array form as $\underline{\hat{\mathbf{Y}}}$), where n_1 is the channel index ranging from 1 to N_1, s_t is the time/segment index ranging from 1 to S_t, f_1 is the frequency index ranging from 1 to F_1, and t_1 is the time index ranging from 1 to T_1. The 4-way PARAFAC is then applied to this 4-D array signal rendering a space-time-frequency-time/segment model (STF-TS model). The STF-TS model of the 4-D array $\underline{\hat{\mathbf{Y}}}$ can be formulated the same way as the STF model as mentioned in Section 2.4.2 except that a new parameter $\hat{\mathbf{D}}$ is added:

$$\underline{\hat{\mathbf{Y}}}_{N_1 \times S_t \times F_1 \times T_1} = f(\hat{\mathbf{A}}, \hat{\mathbf{B}}, \hat{\mathbf{C}}, \hat{\mathbf{D}}) + \underline{\hat{\mathbf{E}}}_{N_1 \times S_t \times F_1 \times T_1}, \tag{4.1}$$

where the 4-way PARAFAC model, i.e. the STF-TS model, is

$$f(\hat{\mathbf{A}}, \hat{\mathbf{B}}, \hat{\mathbf{C}}, \hat{\mathbf{D}}) = \sum_{m=1}^{M} \hat{a}(n_1, m)\hat{b}(s_t, m)\hat{c}(f_1, m)\hat{d}(t_1, m),$$

and $\underline{\hat{\mathbf{E}}}$ is now a 4-D array residual of the model. Each column of $\hat{\mathbf{A}}_{N_1 \times M}$ denotes the space signature of the m-th component ranging from 1 to M where its matrix elements are denoted as $\hat{a}(n_1, m)$. Each column of $\hat{\mathbf{B}}_{S_t \times M}$ denotes the time/segment signature where its matrix elements are denoted as $\hat{b}(s_t, m)$. Each column of $\hat{\mathbf{C}}_{F_1 \times M}$ denotes the

frequency signature where its matrix elements are denoted as $\hat{c}(f_1, m)$. Each column of $\hat{\mathbf{D}}_{T_1 \times M}$ denotes the time signature where its matrix elements are denoted as $\hat{d}(t_1, m)$. Similar to the STF model, the parameters $\hat{\mathbf{A}}, \hat{\mathbf{B}}, \hat{\mathbf{C}}$, and $\hat{\mathbf{D}}$ can be estimated by the ALS with the following cost function

$$\text{argmin}_{\hat{a},\hat{b},\hat{c},\hat{d}} \left\| \underline{\hat{\mathbf{Y}}} - \sum_{m=1}^{M} \hat{a}(n_1, m)\hat{b}(s_t, m)\hat{c}(f_1, m)\hat{d}(t_1, m) \right\|.$$

It should be noted that T in the STF model is equal to $T_1 \times S_t$ in the STF-TS model.

4.2.2 Space-Time-Frequency-Space/Segment Model (STF-SS model)

Instead of segmenting the time domain of the signal, we can also use a similar approach for the space domain. All channels of EEG are first equally divided into groups in the space domain to form a 3-D array signal of space, time, and space/segment domains. Then, a time-frequency transform is applied to each channel to form a 4-D array signal of the space, time, frequency and space/segment domains. After that the 4-way PARAFAC is applied to extract the features of this 4-D array resulting in a space-time-frequency-space/segment model (STF-SS model). The STF-SS model can be mathematically formulated the same way as the STF-TS model except that the selected content to be segmented is in the space domain.

4.2.3 Space-Time-Frequency-Space/Segment-Time/Segment Model (STF-SS-TS model)

If a multi-channel EEG signal happens to have both high number of channels and long period of time, the STF-TS and STF-SS models might not be as useful as they are. Hence, in order to efficiently estimate the STF for this type of signal, the generalization of the STF-TS and STF-SS models called the space-time-frequency-space/segment-time/segment model (STF-SS-TS model) are derived. First, the temporal domain of a multi-channel EEG signal is divided into segments yielding a 4-way array as the input

data of the STF-TS model. After that all channels of the resulting 4-way array are equally divided into groups yielding a 5-way array, $y(s_n, s_t, n_3, t_3, f_3)$ (denoted in array form as $\underline{\mathbf{Y}}$), where s_n is the channel/segment index ranging from 1 to S_n, s_t is the time/segment index ranging from 1 to S_t, n_3 is the channel index ranging from 1 to N_3, f_3 is the frequency index ranging from 1 to F_3, and t_3 is the time index ranging from 1 to T_3. The 5-way PARAFAC is then applied to this 5-D array signal rendering the STF-SS-TS model. The STF-SS-TS model of the 5-D array $\underline{\mathbf{Y}}$ can be formulated by combining the time/segment and space segment/signatures together in one model, that is:

$$\underline{\mathbf{Y}}_{N_3 \times S_t \times F_3 \times T_3 \times S_n} = f(\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}, \mathbf{G}) + \underline{\mathbf{E}}, \tag{4.2}$$

where the 5-way PARAFAC model, i.e. the STF-SS-TS model, is

$$f(\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}, \mathbf{G}) = \sum_{m=1}^{M} a(n_3, m) b(s_t, m) c(f_3, m) d(t_3, m) g(s_n, m),$$

and $\underline{\mathbf{E}}$ is now a 5-D array residual of the size $N_3 \times S_t \times F_3 \times T_3 \times S_n$. Each column of $\mathbf{A}_{N_3 \times M}$ denotes the space signature of the m-th component ranging from 1 to M where its matrix elements are denoted as $a(n_3, m)$. Each column of $\mathbf{B}_{S_t \times M}$ denotes the time/segment signature where its matrix elements are denoted as $b(s_t, m)$. Each column of $\mathbf{C}_{F_3 \times M}$ denotes the frequency signature where its matrix elements are denoted as $c(f_3, m)$. Each column of $\mathbf{D}_{T_3 \times M}$ denotes the time signature where its matrix elements are denoted as $d(t_3, m)$, and each column of $\mathbf{G}_{S_n \times M}$ denotes the space/segment signature where its matrix elements are denoted as $g(s_n, m)$. The parameters $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}$, and $\mathbf{G}$ can be estimated by the ALS where the cost function is

$$\text{argmin}_{a,b,c,d,g} \left\| \underline{\mathbf{Y}} - \sum_{m=1}^{M} a(n_3, m) b(s_t, m) c(f_3, m) d(t_3, m) g(s_n, m) \right\|.$$

It should be noted that T and N in the STF model are equal to $T_3 \times S_t$ and $N_3 \times S_n$ in the STF-SS-TS model, respectively.

4.3 Estimation Methods for Calculating the STF Model from the Reduced Complexity STF Models

In this section, we show that the reduced complexity STF models can be efficiently used for estimating the conventional STF model. Specifically, instead of directly calculating the space, time, and frequency signatures from the original data using the STF model as in Section 2.4, we can estimate these signatures by cascading the weighted versions of their local signatures obtained by the reduced complexity STF models.

4.3.1 Estimation Method for Calculating the STF Model from the STF-TS Model

In this section, we demonstrate a method to estimate the STF model from the STF-TS model, i.e. we have to estimate the signatures of the STF model using the signatures of the STF-TS model. According to (4.1), the time signatures of a long-term multi-channel EEG signal can be estimated by cascading all S_t segments of the time signatures $\hat{\mathbf{D}}$ which are weighted by their corresponding time/segment signatures $\mathbf{B}$. In order to efficiently estimate the STF model from the STF-TS model, the suggested number of segments S_t and number of components M should be the ones that maximize the CORCONDIA value [7]:

$$\text{argmax}_{M,S_t} \left\{ \text{CORCONDIA}(\underline{\hat{\mathbf{Y}}}, \hat{\mathbf{A}}, \hat{\mathbf{B}}, \hat{\mathbf{C}}, \hat{\mathbf{D}}) \right\}. \tag{4.3}$$

In addition, to efficiently estimate the frequency signatures, the length (L) of each segment should satisfy the following condition:

$$L \geq \frac{1}{\min(f_{EEG})}, \tag{4.4}$$

where $\min(f_{\text{EEG}})$ denotes the minimum (fundamental) frequency of the multi-channel EEG signal.

When the residual is neglected, the STF model in (2.4) can also be written in a matrix form as

$$\underline{\acute{\mathbf{Y}}}_{F\times T\times N} = \acute{\mathbf{D}}\left(\Sigma_{\acute{\mathbf{C}}_f}\acute{\mathbf{A}}^T\right), \tag{4.5}$$

where $\Sigma_{\acute{\mathbf{C}}_f}$ is the diagonal matrix with the f-th row of $\acute{\mathbf{C}}$ along the diagonal, $f = 1, 2, ..., F$. With the similar form, the STF-TS model (4.1) can be written in matrix form as

$$\underline{\hat{\mathbf{Y}}}_{F_1\times S_t\times T_1\times N_1} = \hat{\mathbf{D}}\Sigma_{\hat{\mathbf{B}}_{s_t}}\left(\Sigma_{\hat{\mathbf{C}}_{f_1}}\hat{\mathbf{A}}^T\right), \tag{4.6}$$

where $\Sigma_{\hat{\mathbf{C}}_{f_1}}$ is the diagonal matrix with the f_1-th row of $\hat{\mathbf{C}}$ on the diagonal, $f_1 = 1, 2, ..., F_1$, and $\Sigma_{\hat{\mathbf{B}}_{s_t}}$ is the diagonal matrix with the s_t-th row of $\hat{\mathbf{B}}$ along the diagonal, $s_t = 1, 2, ..., S_t$.

According to (4.5) and (4.6), the time signature $\acute{\mathbf{D}}$ of the STF model can be estimated from the signatures of the STF-TS model as

$$\acute{\mathbf{D}} \approx \left(\hat{\mathbf{D}}\boldsymbol{\Sigma}_{\hat{\mathbf{B}}_1}, \ldots, \hat{\mathbf{D}}\boldsymbol{\Sigma}_{\hat{\mathbf{B}}_{S_t}}\right)^T. \tag{4.7}$$

The frequency signature $\acute{\mathbf{C}}$ can be efficiently approximated as $\hat{\mathbf{C}}$ because of the condition in (4.4). The space signature $\acute{\mathbf{A}}$ is approximately equal to $\hat{\mathbf{A}}$. The estimation method for computing the STF model from the STF-SS model can be derived the same way as the STF-TS model except that the segmentation is performed on the space domain.

4.3.2 Estimation Method for Calculating the STF Model from the STF-SS-TS Model

In this section, we aim to estimate the STF model from the STF-SS-TS model. According to (4.2), the time signatures of a multi-channel signal can be estimated by cascading all S_t segments of the time signatures $\mathbf{D}$ which are weighted by their corresponding time/segment signatures $\mathbf{B}$. Similarly, the space signatures of the multi-channels signal can be estimated by cascading all S_n segments of the space signatures $\mathbf{A}$

which are weighted by their corresponding space/segment signatures $\mathbf{G}$. S_t, S_n and M can be selected in order to satisfy the following maximization argument:

$$\text{argmax}_{M,S_t,S_n} \{\text{CORCONDIA}(\underline{\mathbf{Y}}, \mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}, \mathbf{G})\}. \tag{4.8}$$

However, in practice, the calculation of the CORCONDIA value of the STF-SS-TS model may consume high computational resources. In order to solve this problem, we suggest an alternative criterion by maximizing the numbers of iterations by which the ALS algorithm converges. It should be noted that convergence on fitting the model does not necessary imply convergence of parameters, but in practice this is usually the case [50].

When the residual is neglected, the STF-SS-TS model can be written in a matrix form as

$$\underline{\mathbf{Y}}_{F_3 \times S_t \times S_n \times T_3 \times N_3} = \left(\mathbf{D}\Sigma_{\mathbf{B}_{s_t}}\right) \Sigma_{\mathbf{C}_{f_3}} \left(\mathbf{A}\mathbf{\Sigma}_{\mathbf{G}s_n}\right)^T, \tag{4.9}$$

where $\Sigma_{\mathbf{C}_{f_3}}$ is the diagonal matrix with the f_3-th row of $\mathbf{C}$ along the diagonal, $\Sigma_{\mathbf{B}_{s_t}}$ is the diagonal matrix with the s_t-th row of $\mathbf{B}$ along the diagonal, and $\Sigma_{\mathbf{G}_{s_n}}$ is the diagonal matrix with the s_n-th row of $\mathbf{G}$ along the diagonal. $s_t = 1, ..., S_t$, $s_n = 1, ..., S_n$, and $f_3 = 1, ..., F_3$. The time signature $\acute{\mathbf{D}}$ of the STF model can be estimated from the STF-SS-TS model as

$$\acute{\mathbf{D}} \approx \left(\mathbf{D}\mathbf{\Sigma}_{\mathbf{B}_1}, \ldots, \mathbf{D}\mathbf{\Sigma}_{\mathbf{B}_{S_t}}\right)^T. \tag{4.10}$$

Similarly, the space signature $\acute{\mathbf{A}}$ of the STF model can be estimated from the STF-SS-TS model as

$$\acute{\mathbf{A}} \approx \left(\mathbf{A}\mathbf{\Sigma}_{\mathbf{G}_1}, \ldots, \mathbf{A}\mathbf{\Sigma}_{\mathbf{G}_{S_n}}\right)^T. \tag{4.11}$$

Table 4.1. Parameter analysis of the STF model and the reduced complexity STF models

Models	Number of free parameters
STF	$P_{STF} = M(N + F + T)$
STF-TS	$P_{STF-TS} = M(N_1 + S_t + F_1 + T_1)$
STF-SS	$P_{STF-SS} = M(N_2 + F_2 + T_2 + S_n)$
STF-SS-TS	$P_{STF-SS-TS} = M(N_3 + S_t + F_3 + T_3 + S_n)$

4.4 Parameter Analysis

By decomposing the multi-channel EEG signal using the reduced complexity STF models, the number of free parameters [51], i.e. the number of elements that the PARAFAC have to find, can be analyzed in Table 4.1:

- *STF-TS model*: Since T in the STF model is equal to $T_1 \times S_t$ in the STF-TS model, when T is large,

  $$P_{STF-TS} << P_{STF}.$$

  This means that less parameters need to be estimated and thus reduces the computational complexity of the PARAFAC algorithm.

- *STF-SS model*: Given that S_n is the number of segments in a space domain. N_2, F_2, and T_2 are the numbers of channels in one segment, the number of frequency index, and the number of time index, respectively. Since N in the STF model is equal to $N_2 \times S_n$ in the STF-SS model, when N is high,

  $$P_{STF-SS} << P_{STF}.$$

- *STF-SS-TS model*: According to the STF-TS and STF-SS models, it is clear that when T and N are high,

  $$P_{STF-SS-TS} << P_{STF-TS}, P_{STF-SS} << P_{STF}.$$

4.5 Simulation Results

The goal of this section is to investigate the performances among the STF-TS, STF-SS and STF-SS-TS models whether they are good approximations of the STF model for the purpose of real-world applications. We demonstrate the uses of these three models in two applications where the number of channels or the temporal domain of the input signals can be high, e.g. the artifact removal of a multi-channel EEG and the left/right imagery EEG classification.

4.5.1 Eyeblink Artifact Removal

In this experiment, we use a dataset of a 24-channel EEG signal (Fig.4.1). This signal is contaminated by approximately 2Hz eyeblink artifacts in channels 3-10 at the time stems around 0.2, 2.8, 4.2, 7.2, and 8.9 seconds. The goal is to extract these eyeblink artifacts from the 24-channel EEG by using space, time, and frequency information. The conventional STF model and the proposed reduced complexity models are applied to this data.

4.5.1.1 Issue on the Performance

According to the nature of our dataset, the STF model with the number of components (M) equals two (which corresponds to the maximum CORCONDIA value of 96.93) is selected. In order to estimate the STF model, the selected M is then used to further calculate S_t in (4.3). As a result, the maximum CORCONDIA value of 18.04 is achieved when S_t is eighteen. For the STF-SS-TS model, the resulting M and S_t are further used to compute S_n in (4.8). However, as mentioned in Section 4.3.2, the calculation of CORCONDIA values of the STF-SS-TS model is quite complex. Therefore, a convergent criterion is used. The minimum number of iterations used for the STF-SS-TS model of this eyeblink signal is eighteen when S_n is two. In addition, we also investigate the use

of the STF-SS model which employs the resulting S_n from the STF-SS-TS model even though the number of channels of the dataset is not very high.

The space signatures of the STF-TS, STF-SS, and STF-SS-TS models (Figs.4.2(d), (g), and (j), respectively) result in similar signatures with those obtained from the conventional STF model (Fig.4.2(a)). Intuitively, the first component of each model can efficiently extract eyeblink artifacts which mainly occur in channels 3-10. The time signatures of the STF model (Fig.4.2(b)) also contain similar information as the estimated time signatures derived from the STF-TS, STF-SS, and STF-SS-TS models (Figs.4.2(e), (h), and (k), respectively), i.e. the eyeblink artifacts can be distinguished from the background EEG. Even though segmenting the time domain as in the STF-TS and STF-SS-TS models (Fig.4.2(e) and (k)) can cause some distortions in time signatures, the peak locations which are corresponding to all five eyeblink artifacts occurring at times 0.2, 2.8, 4.2, 7.2, and 8.9 seconds can still be preserved. In this experiment, frequency of each eyeblink artifact is approximately 2 Hz. According to Figs.4.2(c), (f), (i), and (l), it is clear that the frequency component of the eyeblink can be well decomposed by the STF model and our reduced complexity models. The STF and STF-SS (Figs.4.2(c) and (i)) models give almost the same signatures, while there are some small distortions in those of the STF-TS and STF-SS-TS (Figs.4.2(f) and (l)) models. This is because segmenting the time domain would cause more effect on changing the fundamental frequency in some intervals than segmenting the space domain.

4.5.1.2 Issue on the Complexity

According to Section 4.2, by using the STF model, we have to calculate the PARAFAC of the 3-way array $\underline{\acute{\mathbf{Y}}}_{N\times F\times T}$ of size $24 \times 91 \times 1800$. This process consumes a longer period of time due to the calculations of more free parameters compared with the STF-TS model in which $\underline{\hat{\mathbf{Y}}}_{N_1\times S_t\times F_1\times T_1}$ is of size $24 \times 18 \times 91 \times 100$. The second and

Table 4.2. Free parameters and normalized time complexity consumed by the STF and STF-TS models of a left eyeblink EEG signal (assume that time consumed by the STF model= 1).

Models	STF	STF-TS	STF-SS	STF-SS-TS
Free parameters	3830	466	3810	446
Time complexity	1	0.121	0.994	0.116
No. of iteration	26	18	28	18

third rows of Table 4.2 illustrate the computational complexities of both the STF and STF-TS models in terms of the numbers of free parameters. By assuming that the computational complexity of the STF model is 1, the STF-TS model consumes only 0.121. It is noted that the free parameters can also be reduced by segmenting the space domain by using the STF-SS model. However, in this experiment, using the STF-SS model is not as efficient as using the STF-TS model since T_1 is much greater than N_2. Further improvement on reducing the computational complexity of the STF and STF-TS models can be done by using the STF-SS-TS model of the 5-way array $\underline{\mathbf{Y}}_{N_3 \times S_t \times F_3 \times T_3 \times S_n}$ of size $2 \times 18 \times 91 \times 100 \times 12$. The STF-SS-TS model consumes 4% less complexity than the STF-TS model and 88.4% less than the STF model. The numbers of iterations used before the ALS converges in order to calculate the free parameters of all the models are also shown in the fourth row of Table 4.2. The results imply that besides the efficiently approximated signatures as in Figs.4.2(d)-(l), all the proposed models also converge as quickly as the conventional STF model.

4.5.2 Left/Right Imagery EEG Signal Classification

In this section, we investigate the uses of the proposed models for extracting features of a multi-channel EEG signal in order to distinguish between left and right imagery

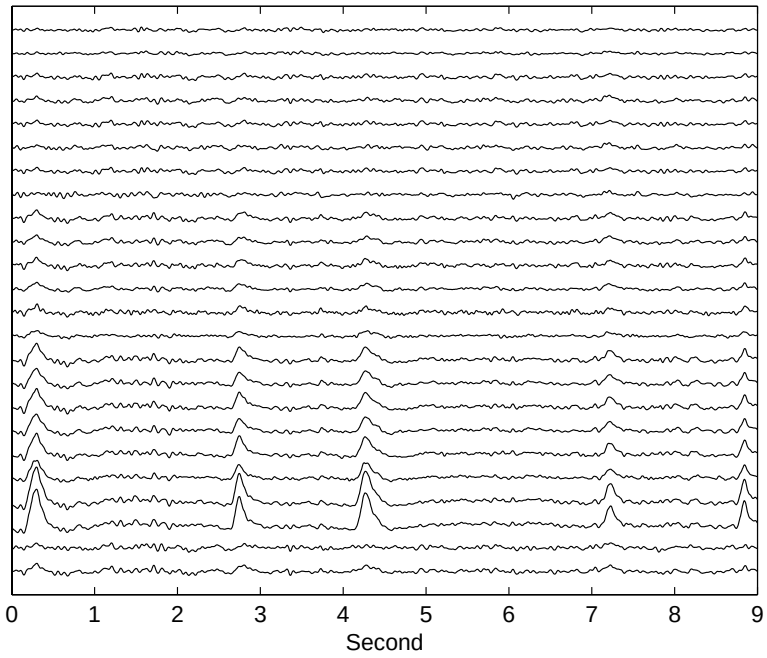

Figure 4.1. Original eyeblink artifact data of 24-channel EEG: Lowest and highest lines represent channels 1 and 24, respectively.

signals. The dataset in [52] which contains nine subjects of 59-channel EEG at the sampling rate of 100 Hz is used. Each subject is asked to perform some specific tasks, e.g. push imagined left or right bottom. For simplicity, 4.16 second of 24-channel EEG is selected from the original 6 second of the 59-channel EEG. Ninety trials of the 24-channel EEG are used for training, and the rest ninety trials are used for testing. Subjects 3, 6, 8 and 9 are used in this experiment.

Originally, the STF model for this left/right imagery signal classification is used in [53]. In [53], a multi-channel EEG signal is decomposed using a 3-way PARAFAC, then the space signature of the selected component, i.e. the component of PARAFAC selected from all M components, is employed as the feature vector for classification using the support vector machine (SVM). To compare the performances of the STF, STF-SS

Table 4.3. Classification accuracy (%) for left/right imagery signal classification of 4 subjects using space (S), estimated space (ES) and space/segment (SS) signatures as feature vectors (FV).

Models	STF	STF-SS	STF-SS	STF-SS-TS	STF-SS-TS
Signatures	S	ES	SS	ES	SS
Lengths of FV	24	24	12	24	12
Subject 3	60	57.78	63.33	55.56	61.11
Subject 6	55.56	60	54.44	55.56	61.11
Subject 8	54.44	57.78	57.78	57.78	57.78
Subject 9	56.67	54.44	58.89	62.22	57.78

and STF-SS-TS models, we follow the same process as in [53] except that, instead of using the SVM, only simple linear discriminant analysis (LDA) is used for our classification experiment.

In this experiment, the space domain is divided into twelve groups (segments), and each group contains 2-channel EEG as follows:

- Segment 1 (channels 1-2): FT7, T7,
- Segment 2 (channels 3-4): TP7, CP5,
- Segment 3 (channels 5-6): C5, FC5,
- Segment 4 (channels 7-8): FC3, C3,
- Segment 5 (channels 9-10): CP3, CP1,
- Segment 6 (channels 11-12): C1, FC1,
- Segment 7 (channels 13-14): FC2, C2,
- Segment 8 (channels 15-16): CP2, CP4,
- Segment 9 (channels 17-18): C4, FC4,
- Segment 10 (channels 19-20): FC6, C6,
- Segment 11 (channels 21-22): CP6, TP8,

Table 4.4. Classification accuracy (%) for left-right imagery signal classification of 4 subjects using space/segment (SS) signatures, time/segment (TS) signatures, and both space/segment signatures and time/segment signatures as the feature vectors (FV).

Models	STF-SS	STF-SS-TS	STF-TS	STF-SS-TS	STF-SS-TS
Signatures	SS	SS	TS	TS	SS+TS
Lengths of FV	12	12	8	8	20
Subject3	63.33	61.11	57.78	56.67	62.22
Subject6	54.44	61.11	53.33	54.44	62.22
Subject8	57.78	57.78	60	62.22	63.33
Subject9	58.89	57.78	54.44	61.11	60

Table 4.5. Free parameters, their corresponding complexity and the average numbers of iterations (used before the ALS converges) consumed by the STF, STF-TS, STF-SS and STF-SS-TS models of the left/right imagery EEG signal (assume that the complexity consumed by the STF model= 1).

Models	STF	STF-TS	STF-SS	STF-SS-TS
Free parameters	1122	410	1102	390
Complexity	1	0.365	0.982	0.348
Avg. no. of iterations	22.69	27.69	28.06	24.56

- Segment 12 (channels 23-24): T8, FT8.

The second, third and fifth columns of Table 4.3 demonstrate that the classification accuracies obtained by using the signatures of the STF-SS model ($S_n = 12,\ M = 2$) and STF-SS-TS model ($S_n = 12,\ S_t = 8,\ M = 2$) as feature vectors are comparable, and they are also comparable with the classification accuracies obtained from the STF model. In addition, the estimated space signatures using the STF-SS and STF-SS-TS models are reconstructed from cascading the weighted versions of the space signatures with their corresponding space/segment signatures. Therefore, in this classification experiment, it makes sense to employ only the space/segment signatures of the STF-SS and STF-SS-TS

models as feature vectors since they have much shorter lengths. The fourth and sixth columns of Table 4.3 show that using length-12 space/segment signatures as feature vectors also yields comparable classification accuracies with the STF model while the lengths of feature vectors are reduced by a factor of two. Taking into account that the event-related desynchronization and event-related synchronization (ERD/ERS) patterns which are the keys to distinguish between left and right imagery signals might take place only on channels C3, C4 and some of their neighboring channels [54]. Therefore, including too many channels, e.g. 24 channels, might degrade the classification accuracy. However, it is suitable to see the performance of our reduced complexity models.

To improve the classification accuracy, we include the time/segment signatures as additional features under two main assumptions:

- Besides the space domain, the ERD/ERS patterns can also be observed in the time domain.
- By using fewer features, time/segment signatures of the STF-TS and STF-SS-TS models contain similar information to the time signatures of the STF model,

Table 4.4 shows improvement in classification accuracy over the methods used in Table 4.3. This implies that both space and time domains of the EEG signals are useful features for EEG classification and can be simultaneously and efficiently extracted by using the STF-SS-TS model. Table 4.5 further illustrates that while the numbers of iterations (used before the ALS converges) are comparable, the computational complexity for calculating the free parameters of the STF-SS, STS-TS and STF-SS-TS models is reduced from the STF model by 1.8%, 63.5% and 65.2%, respectively.

4.5.3 Removal of the Eyeblink Artifact from EEGs Using the STF-TS Model and Robust Minimum Variance Beamforming (RMVB)

In EEG analysis, clean EEG signals without any artifact, e.g. eyeblink, speaking, ECG, are usually preferred by the doctor. Therefore, in this experiment, we aim to

recover the clean signals back from the eyeblink-contaminated signals by a more efficient and sophisticated method than the decomposition in Section 4.5.1.

Let $\mathbf{s}(t) = [s_1(t), s_2(t), \cdots, s_N(t)]^T$ be N zero-mean real mutually uncorrelated point geometrically stationary sources which are mixed by an $N \times N$ full column rank matrix (mixing matrix) $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_N]$ where $\mathbf{a}_i$ is the i-th column of $\mathbf{A}$. The vector of time mixture samples $\mathbf{x}(t) = [x_1(t), x_2(t), \cdots, x_N(t)]^T$ which is shown in Fig.4.3 can be formulated as

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) + \mathbf{v}(t) \tag{4.12}$$

where $\mathbf{v}(t) = [v_1(t), v_2(t), \cdots, v_N(t)]^T$ is the additive white Gaussian zero-mean noise which is assumed to be spatially uncorrelated with the sensor data and temporally uncorrelated.

4.5.3.1 Implementation of the RMVB

According to Fig.4.3, since $\mathbf{a}_j$ performs as the steering vector of the j-th source, we may write $y(t)$ which is an estimation of the source $s_j(t)$ as

$$y(t) = \mathbf{w}_j^T \mathbf{x}(t),$$

where $\mathbf{w}_j$ acts as a spatial filter. $\mathbf{w}_j$ can be found by minimizing the energy of $y(t)$ subject to the prior knowledge on how the eyeblink artifacts look like. Since the prior knowledge we use is not the ideal one, $\mathbf{w}_j$ can be found by solving the worst-case performance optimization [55] where the estimated prior knowledge can be used. In general, we can write this problem as

$$\text{Minimize} \quad J_c = \mathbf{w}_j^T \mathbf{R} \mathbf{w}_j \quad \text{subject to} \quad \min_{\|\delta\| \leq \epsilon} |\mathbf{w}_j^T \hat{\mathbf{a}}_j + \mathbf{w}_j^T \delta| = 1,$$

where $\mathbf{R} = \frac{1}{K}\sum_{k=1}^{K}\mathbf{R}_{\mathbf{xx}}^{k}$, and $\delta = \mathbf{a}_j - \hat{\mathbf{a}}_j$. By using the Lagrange multiplier method, differentiating J_c with respect to $\mathbf{w}_j$, and setting to zero, the spatial filter $\mathbf{w}_j$ can be computed as follows:

$$\mathbf{w}_j = \left[\mathbf{R} + \frac{\epsilon}{\rho}\mathbf{I}\right]^{-1}\hat{\mathbf{a}}_j, \tag{4.13}$$

where $\rho \equiv \|\mathbf{w}_j\|$ and $\mathbf{I}$ denotes the identity matrix. This equation can be simplified by using the eigendecomposition of $\mathbf{R}$, i.e.

$$\mathbf{R} = \mathbf{U}\mathbf{\Sigma}\mathbf{U}^T, \tag{4.14}$$

where Σ is an $N \times N$ diagonal matrix whose the diagonal elements are the sorted (in the decreasing order) eigenvalues ($\sigma_i,\ i = 1, ..., N$), and $\mathbf{U}$ is an $N \times N$ matrix whose columns are the eigenvectors of $\mathbf{R}$. With some manipulations on (4.13) [56], we can use ρ that satisfies the following equation

$$\sum_{i=1}^{N}\left(\frac{|g_i|}{\epsilon + \rho\sigma_i}\right)^2 - 1 = 0, \tag{4.15}$$

where $\mathbf{g} = [g_1, g_2, ..., g_N]^T = \mathbf{U}^T\hat{\mathbf{a}}_j$.

The clean multi-channel EEG without the eyeblink contamination, $\mathbf{x}_{filt}(t)$, can be obtained by deflation method [41], that is

$$\mathbf{x}_{filt}(t) = \mathbf{x}(t) - \tilde{\mathbf{w}}_j y(t),$$

where $\tilde{\mathbf{w}}_j$ can be found so that $\mathbf{x}_{filt}(t)$ has minimum energy when the eyeblink source $y(t)$ is removed from the contaminated source $\mathbf{x}(t)$. Specifically, $\tilde{\mathbf{w}}_j$ can be obtained by minimizing $J_j(\tilde{\mathbf{w}}_j)$ with repect to $\tilde{\mathbf{w}}_j$ where

$$\begin{aligned} J_j(\tilde{\mathbf{w}}_j) &= E[\mathbf{x}_{filt}^T(t)\mathbf{x}_{filt}] \\ &= E[\mathbf{x}_j^T(t)\mathbf{x}_j(t)] - 2\tilde{\mathbf{w}}_j^T E[\mathbf{x}_j(t)y(t)] + \tilde{\mathbf{w}}_j^T\tilde{\mathbf{w}}_j E[y(t)^2]. \end{aligned}$$

Hence,

$$\tilde{\mathbf{w}}_j = \frac{E[\mathbf{x}(t)y^T(t)]}{E[y(t)^2]} = \frac{E[\mathbf{x}(t)\mathbf{x}^T(t)]\mathbf{w}_j}{E[y(t)^2]}. \tag{4.16}$$

4.5.3.2 Simulation Results

We applied the artifact removal algorithm in Section 4.5.3.1 to real EEG measurements. The database was provided by the School of Psychology, Cardiff University, UK, and represent a wide range of eyeblinks, i.e. more than 500 eyeblink-contaminated EEG recordings. The scalp EEG was obtained using 25 Silver/Silver-Chloride electrodes placed at locations defined by the conventional 10-20 system [1]. The data were sampled at 200 Hz, and bandpass filtered with cut-off frequencies of 1 Hz and 30 Hz. The performance of the algorithm can be observed by comparing the EEGs obtained at the electrodes in the left subplot of Fig. 4.4 and the same segment of data after being processed by the proposed algorithm in the right subplot of Fig. 4.4.

In order to provide a quantitative measure of performance for the artifact removal method in Section 4.5.3.1, the correlation coefficient (CC) between the extracted eyeblink artifact source, and the original EEGs and the artifact removed EEGs are computed and demonstrated in Fig. 4.5. The CC of two discrete random variables x and y over a fixed interval is mathematically defined as:

$$CC = \frac{|\sum_{i=1}^{w} x(i)y(i)|}{\sqrt{\sum_{j=1}^{w} x^2(j)}\sqrt{\sum_{j=1}^{w} y^2(j)}} \tag{4.17}$$

where w is the number of time samples.

The values reported in Fig. 4.5 have been computed as follows. For each of the 20 different eyeblink-contaminated EEGs, we executed method in Section 4.5.3.1. The aforementioned CCs for each run were then computed between the extracted eyeblink and the EEGs before and after the artifact removal. These values have subsequently been averaged and shown in Fig. 4.5. Furthermore, their corresponding standard deviations have also been reported. The CC values have been significantly decreased by using our artifact removal method. Simulations for 20 EEG measurements demonstrate that the method in Section 4.5.3.1 can efficiently identify and remove the eye-blink artifact from

the raw EEG measurements by using the space signature (according to the eyeblink) of the STF-TS model as a prior knowledge.

4.6 Summary

We have presented three reduced complexity STF models named the STF-TS, STF-SS and STF-SS-TS models. These proposed models can exploit the space, time, frequency, space/segment and time/segment domains of a multi-channel EEG. We also derive the formulae for estimating the STF model from these reduced complexity models. Since, the efficiency of the reduced complexity models depend closely on the choices of selected numbers of segments, we present a criterion based on maximizing the CORCONDIA value in order to obtain the efficient choices of numbers of segments. Besides maximizing the CORCONDIA value, a fast method for finding the number of segments by minimizing the number of iterations before the model converges is also presented. With less computational complexity, the proposed models can efficiently extract the eyeblink artifacts from the normal multi-channel EEG. The proposed models also yield comparable classification accuracies to the conventional STF model when being applied to left/right imagery EEG classification. Finally, the space signature of the STF-TS model can be efficiently used as a prior knowledge to recover back the clean EEG signals from the eyeblink-contaminated signals via the RMVB.

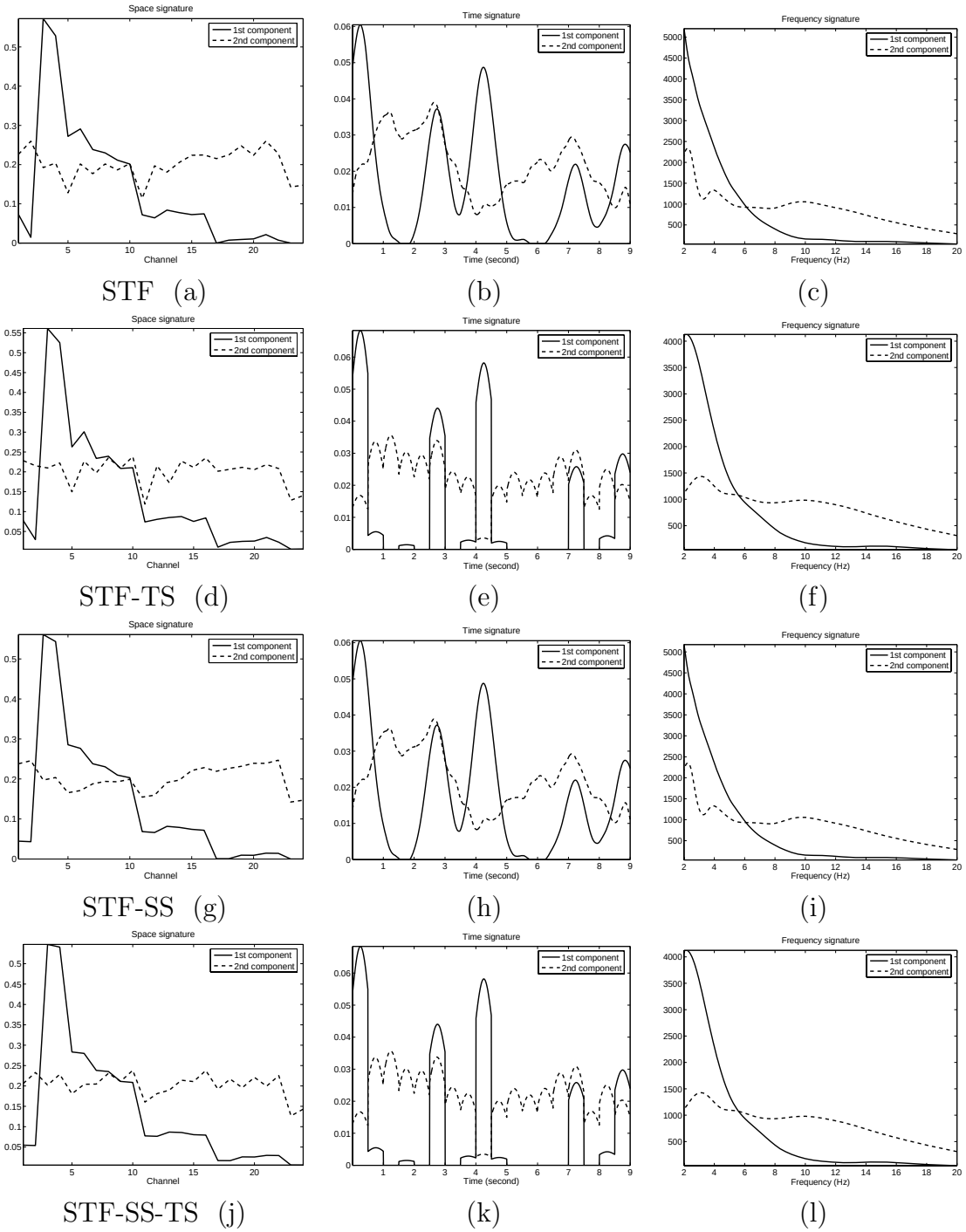

Figure 4.2. Space signatures of the (a) STF, (d) STF-TS, (g) STF-SS and (j) STF-SS-TS models. Time signatures of the (b) STF, (e) STF-TS, (h) STF-SS and (k) STF-SS-TS models. Frequency signatures of the (c) STF, (f) STF-TS, (i) STF-SS and (l) STF-SS-TS models.

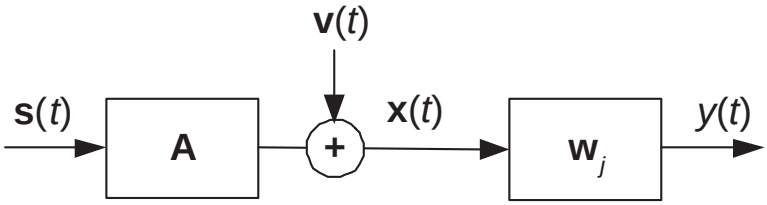

Figure 4.3. System on extracting an eyeblink component from a multi-channel EEG.

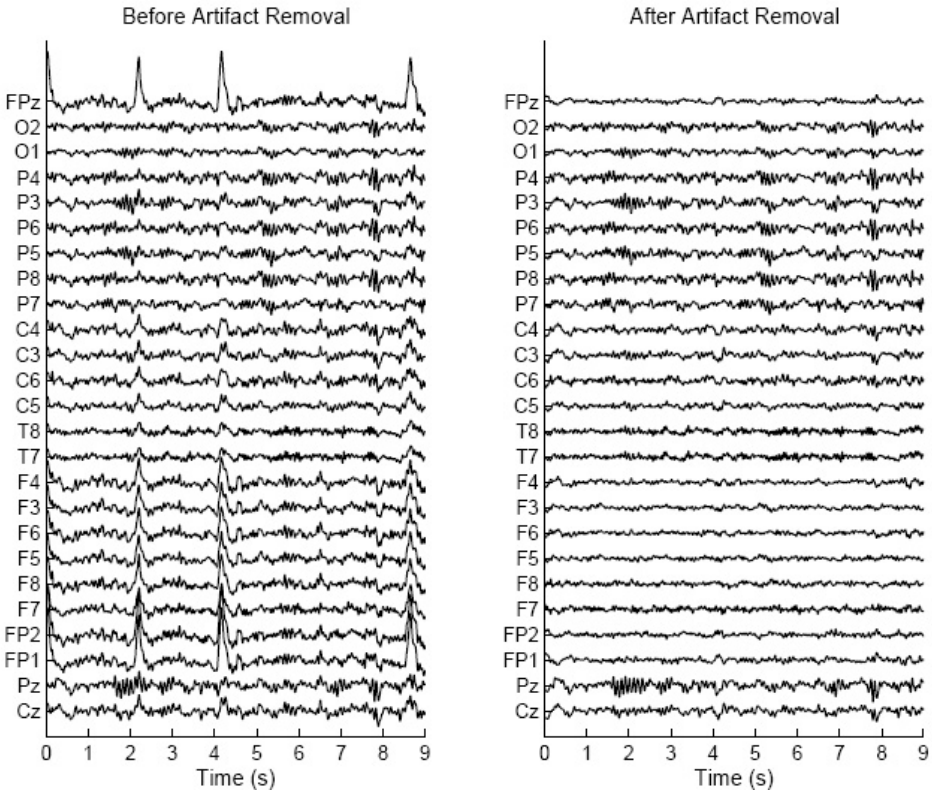

Figure 4.4. The results of the proposed eyeblink artifact removal method for a set of real EEGs. The left subplot depicts highly eyeblink-contaminated EEGs before artifact removal while in the right subplot the segment of EEGs after being corrected for eyeblink artifact is illustrated.

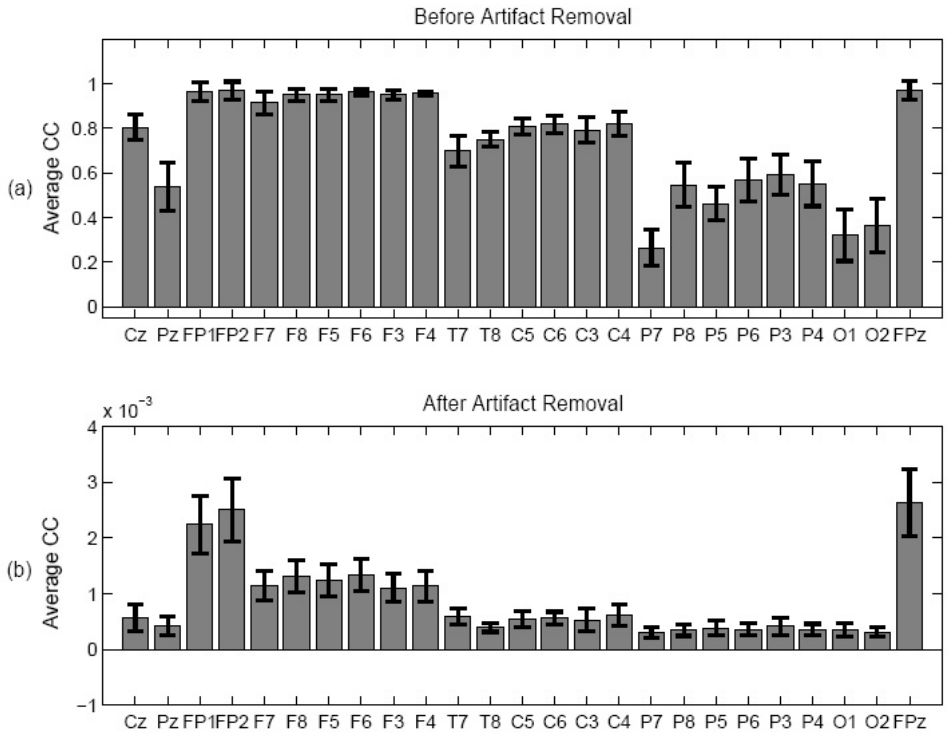

Figure 4.5. The averaged CC values and their corresponding standard deviations between the extracted eyeblink and the restored EEGs (a) before and (b) after artifact removal of different channels. The experiments have been performed for 20 different eyeblink-contaminated EEG recordings.

CHAPTER 5
MULTI-CHANNEL FLEXIBLE LOCAL DISCRIMINANT BASES FOR LEFT/RIGHT IMAGERY EEG CLASSIFICATION

5.1 Introduction

Nowadays, a study on the BCI raises a lot of signal processing issues to be solved [35]. One of that is how to distinguish between the left and right imagery signals and correctly classify them. The possible applications of this issue are such as controlling the wheelchairs, mouse, or remote by using brain signals, e.g. EEG.

Distinguishing the left and right imagery EEG is possible since the event-related desynchronization and synchronization (ERD/ERS) patterns usually occur on the opposite sides of the imagination of a hand movement [54]. This observation on the ERD/ERS patterns motivates many researchers to explore novel theories and algorithms for left/right imagery EEG classification. In [57], a time-frequency based approach is proposed by filtering the fixed time windows in order to obtain band powers (BP) and classifying the resulting BP with the learning vector quantization (LVQ). Automated approach to adjust the influence of the BP during the learning process can be done using the distinction-sensitive learning vector quantization (DLVQ) instead of the LVQ [58]. An alternative way to obtain useful features for the classification is by employing parameters of the autoregressive (AR) model over uniformly short intervals [59], [60]. To further improve [59] and [60], AR parameters are designed to be time dependent by using the model called adaptive autoregressive (AAR) [61]. Taking into consideration that the features in [58]-[61] are designed based on fixed time segments. Since the ERD/ERS patterns might not uniformly occur in time, the classification accuracy might be degraded in some cases. This problem can be efficiently solved by extracting the features of channels C3 and C4

of a multi-channel EEG based on the local discriminant bases (LDB) procedure derived from the local cosine packets (LCP) [11] over nonuniform segments [3],[62]. However, according to the studies in [54], the ERD/ERS patterns at different frequencies might not occur at the same channel. Since only channels C3 and C4 are used in [3],[62], the misclassification rates might be high in some sets of data. In [63], by incorporating more channels besides C3 and C4, the multi-channel classification scheme is illustrated but still yields insignificant improvements compare with those in [3]. For convenience, the classification scheme in [3] is called the conventional scheme.

This thesis aims to improve the LDB in [3] by including more useful channels rendering a more efficient feature extraction scheme called multi-channel flexible local discriminant bases (MF-LDB). We also propose two methods to design the MF-LDB. For the first method called hard decision making (HDM), the MF-LDB is designed based on the channel that maximizes class separability. For the second method called soft decision making (SDM), the MF-LDB is designed based on linear combinations of the $\hat{M}$ channels which have the highest class separability. Since, the LCP used for designing the MF-LDB is not a shift invariant transform, spin cycle procedure [64] is employed prior to the design process. After that the important features are selected using Fisher class separability criterion and classified using the linear discriminant analysis (LDA) [65]. To further improve the classification accuracy, we propose a high resolution transform which motivated from the LCP and minimum variance distortionless response [12],[66] called local MVDR packets. The local MVDR packets transform is used instead of the LCP to design the LDB of the conventional scheme and yields approximately 4% improvement in classification accuracy. The publication related to this chapter can be found in [67] and [68].

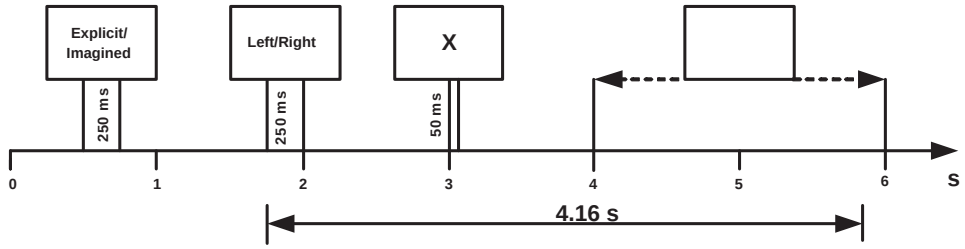

Figure 5.1. Data acquisition stage of the EEG signal used for classification.

5.2 Data Acquisition

The dataset used in this chapter is obtained from the 2002 BCI competition [52]. This dataset contains nine subjects of 59-channel EEG at a sampling rate of 100 Hz. Each subject is asked to perform some specific tasks, e.g. push imagined left or right bottom. The experiment lasts for six seconds for each trial and 4.16-second signals after the appearance of the left/right cues are used for the classification (Fig. 5.1). The Hjort derivation [69] is applied to each channel, $\mathbf{e}_i$, in order to obtain local activities. The resulting channel after performing the Hjort derivation, $\mathbf{e}_i^{Hjort}$ can be approximated as follows:

$$\mathbf{e}_i^{Hjort} = \mathbf{e}_i - \frac{1}{4}\sum_{j \in N_i} \mathbf{e}_j, \tag{5.1}$$

where N_i denotes the indices of four neighboring channels of $\mathbf{e}_i$. The EEG signals are then filtered to retain frequencies between 2-40 Hz which are found to be a meaningful frequency range used in the analysis of EEG [35], [54]. In this paper, all one hundred and eighty trials (ninety trials of left and right imagery signals) are used for the classification.

5.3 Conventional EEG Classification Method

This section reviews the conventional method for left/right imagery classification of EEG signals proposed in [3]. The block diagram of the conventional classification methods is shown in Fig. 5.2(a). The flexible local discriminant bases (F-LDB) designed over the

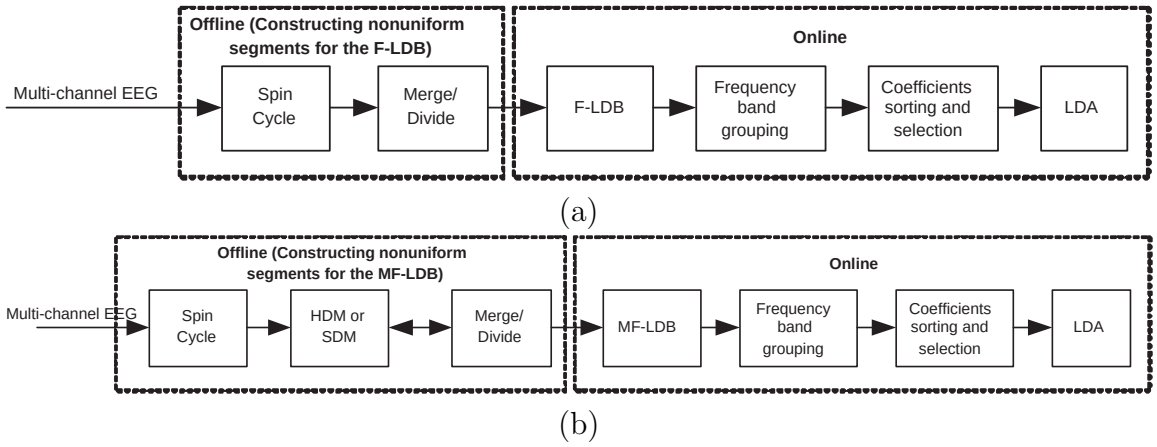

Figure 5.2. Block diagram of (a) the conventional left/right imagery classification method [3] and (b) the proposed classification methods.

nonuniform segments derived from the LCP is used as the feature extraction procedure. Normally, some other modalities can be used as local bases, e.g. wavelet packets. The LCP is selected as the local bases since ERD/ERS patterns can be designed as a time locked event-related potentials [62]. To avoid the shift varying issue of the LCP, spin cycle procedure [64] is employed. It is noted that the design process of the F-LDB can be implemented offline. For the online process, the designed F-LDB is used to extract the features. The resulting LCP coefficients are then sorted according to their class separability. The linear discriminant analysis (LDA) is used as the classifier.

5.3.1 Flexible Local Discriminant Bases (F-LDB)

The design of the F-LDB is composed of two main components: 1) *spin cycle procedure* which is used to eliminate the shift varying issue of the LCP, and 2) *merge and divide procedure* which is used to design efficient nonuniform segments for the F-LDB.

5.3.1.1 Spin Cycle Procedure

Taking into consideration that the ERD/ERS patterns of each trial may not exactly occur at the same time, therefore the transform used for these types of input should be shift invariant. Since the LCP is not a shift invariant transform, i.e. LCP coefficients of the original data and their shifted versions are different, we may not obtain the efficient segments for constructing the LDB. This problem can be alleviated by using the spin cycle procedure [64], i.e. we also include the shifted versions (by $-\tau, ..., \tau$) of the original signal for constructing the nonuniform segment of the LCP. In particular, we have 2τ more input signals for constructing the F-LDB via the LCP.

5.3.1.2 Merge and Divide Procedure

Merge and divide procedure is used to find the optimal nonuniform segments for the F-LDB which result in features that maximize class separability. This procedure can be summarized as follows:

1. Divide the EEG signals into small uniform segments (up to the required frequency resolution), calculate their LCP coefficients and construct children and mother structures as in Fig. 5.3(a), e.g. $\mathbf{M}_1$ is a mother segment with the corresponding two children segments $\mathbf{y}_1$ and $\mathbf{y}_2$,
2. For each mother segment and its corresponding two children segments, calculate the Euclidean distances of the cumulative distribution functions (cdf-distances) [64] of their LCP coefficients as the class separability,
3. Merge the children segments if the sum of their distances is less than the distance of their corresponding mother segment; otherwise divide the signal at that point,
4. Continue the previous step from left segments to right segments until nonuniform segments are obtained.

It should be noted that calculating the distance in the transform domain, which is quite sensitive to the outliers in the dataset, sometimes may not lead to the true discriminant patterns. This problem can be alleviated by computing the distance of the cdf of the LCP coefficients [64].

5.3.2 Feature Extraction and Dimension Reduction

Once the nonuniform segments are obtained from the merge and divide procedure, their corresponding LCP coefficients are used as the features for the classification. However, we cannot use all features we have because of the curse of dimensionality [70]. Therefore, we group the LCP coefficients of channels C3 and C4 together and sort them using the Fisher class separability criterion:

$$F = \frac{(\mu_1 - \mu_2)^2}{\sigma_1^2 - \sigma_2^2}, \tag{5.2}$$

where μ_i and σ_i^2 are mean and variance of the feature vector of class i. The top k LCP coefficients will be further used as the selected features for classification.

5.3.3 Classification

For simplicity, linear discriminant analysis (LDA) is employed where the distance of the features for discriminating the hyperplane is

$$d = \mathbf{v}^T \mathbf{s} - v_0,$$

where $\mathbf{s}$ is the feature vector, v_0 is the threshold for making a decision, and $\mathbf{v}$ is the weight vector of the LDA which can be computed by

$$\mathbf{v} = \left(\sum\nolimits_1 + \sum\nolimits_2 \right)^{-1} (\mathbf{m}_1 - \mathbf{m}_2),$$

where $\sum_i$ and $\mathbf{m}_i$ are the covariance matrix and the mean vector of the feature vectors of class i.

5.4 Multi-Channel Flexible Local Discriminant Bases (MF-LDB)

Unlike the F-LDB, we propose the local discriminant bases which also take into account channels besides C3 and C4 called multi-channel flexible local discriminant bases (MF-LDB). In other words, channels of interest COI_{C3} in Fig. 5.3(c) and COI_{C4} in Fig. 5.3(d) are used to design the MF-LDB. Based on the problem formulation, we suggest two methods to design the MF-LDB named hard decision making (HDM) and soft decision making (SDM). For simplicity, let us consider using only the COI_{C3} for describing our decision making methods in Sections 5.4.1-5.4.3.

5.4.1 Problem Formulation

Since the ERD/ERS patterns of each frequency band of EEG usually occur at different scalp locations, using more than just C3 and C4 electrodes may improve the classification performance in some situations. Thus, the MF-LDB which is an adaptive method that takes into account multi-channel EEG can be formulated as follows:

Let the $M \times L$ matrix $\mathbf{X}_n$ be the n-th segment of a length-L M-channel signal (Fig.5.3(b)), and let the length-M vector $\mathbf{a}_n$ be the corresponding weight vector of the n-th segment of M-channel signal where $n = 1, ..., N$, N is the number of all children segments (according to the merge and divide procedure), and M is the number of channels of interest. According to Fig.5.3(c), M is nine. Furthermore, let a length-L vector $\mathbf{y}_n$ be a children segment (of the n-th segment) of the merge and divide procedure, then

$$\mathbf{y}_n = \mathbf{X}_n^T \mathbf{a}_n. \tag{5.3}$$

In this section, we describe two efficient methods to find $\mathbf{a}_n$: HDM and SDM.

5.4.2 Hard Decision Making Method (HDM)

In this method, the MF-LDB is designed based on a channel that maximizes the cdf-distance of LCP coefficients between two classes (left and right imagery signals). This method is called hard decision making (HDM).

For the n-th segment, $a_n(m)$ (the element of $\mathbf{a}_n$ corresponding to channel-m) can be set to one when m denotes the channel that maximizes the cdf-distance between two classes, otherwise $a_n(m)$ is zero. However, given the prior knowledge that C3 is known as the significant channel in which ERD/ERS patterns usually clearly occur, we can improve the decision making by giving more priority to this particular channel, i.e. the threshold α needs to be included in the decision making method. Specifically, let $m \in \mathrm{COI}_{C3}$,

1) if m is channel C3,

$$a_n(m) = \begin{cases} 1, & \text{if} \quad \|p_n^m - q_n^m\|^2 \geq \alpha \left\|p_n^{\hat{m}} - q_n^{\hat{m}}\right\|^2, \\ 0, & \text{otherwise,} \end{cases}$$

2) if m is not channel C3,

$$a_n(m) = \begin{cases} 1, & \text{if} \quad \alpha \|p_n^m - q_n^m\|^2 \geq \alpha \left\|p_n^{\hat{m}} - q_n^{\hat{m}}\right\|^2 \geq \left\|p_n^{C3} - q_n^{C3}\right\|^2, \\ 0, & \text{otherwise,} \end{cases}$$

where $\hat{m} \in \mathrm{COI}_{C3} \backslash \{\mathrm{C3}\}$, and p_n^j and q_n^j denote the cdf-distances of LCP coefficients of the n-th segment of classes 1 and 2 data of channel-j, respectively. The threshold α is chosen between zero and one.

5.4.3 Soft Decision Making Method (SDM)

Basically, EEG signals can be modeled as the sum of neural potentials. By assuming that the clear ERD/ERS patterns might not exactly occur at any position of COI_{C3}, a soft decision making method (SDM) is proposed to construct the signal $\mathbf{y}_n$ at a new

position that maximizes the cdf-distance by linearly interpolating the channels selected from COI_{C3}. Hence, $a_n(m)$ can be considered as the weight of channel-m of the n-th segment used for linear interpolation. Similar to the HDM, given a prior knowledge that C3 is known as a significant channel in which ERD/ERS patterns usually occur, the threshold β (ranging from zero to one) also needs to be employed in the decision making method. Specifically, let the set of the first $\hat{M}$ channels that maximize the cdf-distance be $\overline{\mathrm{COI}}_{C3}$ where $\overline{\mathrm{COI}}_{C3} \subseteq \mathrm{COI}_{C3}$. A new channel, say $\tilde{m}$, is constructed by linear interpolation among all $m \in \overline{\mathrm{COI}}_{C3}$ by using $a_n(m)$ as their corresponding weights:

$$a_n(m) = \frac{\|p_n^m - q_n^m\|^2}{\sum_{i \in \overline{\mathrm{COI}}_{C3}} \|p_n^i - q_n^i\|^2}. \tag{5.4}$$

After that two conditions need to be checked in order to obtain $\mathbf{y}_n$.

1) If $\beta \left\|p_n^{\tilde{m}} - q_n^{\tilde{m}}\right\|^2 \geq \left\|p_n^{C3} - q_n^{C3}\right\|^2$, $\mathbf{y}_n$ can be obtained by (5.3) using the resulting $\mathbf{a}_n$ from (5.4).

2) If $\beta \left\|p_n^{\tilde{m}} - q_n^{\tilde{m}}\right\|^2 < \left\|p_n^{C3} - q_n^{C3}\right\|^2$, channel C3 is selected as $\mathbf{y}_n$ without performing any interpolation.

Similarly, the HDM and SDM are also separately applied to COI_{C4} in Fig.5.3(d). The resulting children segments from the HDM or SDM are used in the merge and divide procedure (Section 5.3.1). Once the nonuniform segments and their corresponding LCP coefficients are obtained, we group each frequency bin of the LCP coefficients obtained from both COI_{C3} and COI_{C4} together and sort them using the Fisher class separability criterion in (5.2). The top k coefficients will be further used as the selected features for classification. It is noted that the process of designing the MF-LDB can be implemented offline, hence the complexity of the online process (feature extraction, dimension reduction and classification) is the same as that of the conventional method.

5.4.4 Summary of the Classification Methods Using the HDM and SDM-based MF-LDB

The block diagram of the proposed classification methods is shown in Fig. 5.2(b). It can be summarized as follows:

1. Construct nonuniform segments for the MF-LDB from both COI_{C3} and COI_{C4} using the HDM or SDM together with the merge and divide procedure as an offline process,
2. As an online process, calculate the LCP coefficients according to the resulting MF-LDB from 1),
3. Group the LCP coefficients into selected frequency bins. (Normally, nonuniform frequency bins can be designed offline so that the class separability is maximized, but, experimentally, it yields slightly improvement in classification accuracies over using the uniform ones [62]. Since the ERD/ERS patterns are proved to be mostly occurred in the mu and beta bands [54], hence, in this chapter, only two fixed frequency bins of 8-12 and 16-20 Hz which correspond to the μ and β bands, respectively, are employed).
4. Sort the resulting coefficients from 3. by Fisher class separability (Section 5.3.2 and select the top k coefficients as the selected features for classification,
5. Perform classification using the LDA with 10-fold cross -validation.

As mentioned, since the LCP is not a shift invariant transform, we also employ the spin cycle procedure, i.e. we also include the shifted versions (by $-\tau, ..., \tau$) of the signals in COI_{C3} and COI_{C4} for designing the MF-LDB.

5.4.5 Local MVDR Packets

In [3], the LCP is found to be useful for the left/right imagery EEG classification. Since the LCP is one type of the time-frequency nonredundant transform which is inde-

pendent from the data, it is interesting to develop a transform with similar properties but can be designed based on the input data. According to Section 2.6, by calculating the spectrum of each nonoverlapping segment using the minimum variance distortionless response (MVDR) [12], [66], a new time-frequency nonredundant transform which takes into account information of the input data is proposed called local MVDR packets. Specifically, let $x(t)$ be a real discrete time signal of length $\hat{T}$ where $t = 0, ..., \hat{T} - 1$, the local MVDR packets coefficient at a translation index k and a frequency index ω can be obtained by

$$S(k, \omega) = \frac{1}{\mathbf{e}^H(\omega)\mathbf{R}_{xx}^{-1}(k)\mathbf{e}(\omega)}, \tag{5.5}$$

where $\mathbf{e}(\omega) = [1, e^{-j\omega}, ..., e^{-j\omega(p-1)}]^T$, H denotes a conjugate transpose operator, p denotes length of a local basis of the MVDR (length of the filter), and the autocorrelation matrix $\mathbf{R}_{xx}(k)$ of size $p \times p$ can be estimated as

$$\mathbf{R}_{xx}(k) = \sum_{t=n_k+p}^{n_{k+1}} \left[x(t-1) \ ... \ x(t-p)\right]^T \left[x(t-1) \ ... \ x(t-p)\right],$$

where $0 \leq k \leq K-1$, n_k denotes the time index where the segmentation is performed, n_0 is 0, and n_K is $\hat{T}$. In practice, in order to avoid from being singular, $\mathbf{R}_{xx}(k)$ needs to be added by a matrix $\epsilon\mathbf{I}$ where $\mathbf{I}$ is the $p \times p$ identity matrix and ϵ is a very small positive number. The MF-LDB can also be obtained from the local MVDR packets by using the HDM or SDM followed by the merge and divide procedure.

5.5 Simulation Results

5.5.1 Design Example of HDM and SDM

Let us consider subject nine in the experiment. Channel C4 and its neighboring channels (Fig. 5.3(d)) are used to illustrate the uses of HDM and SDM for designing the MF-LDB. The resulting children segments obtained from the HDM (using threshold

$\alpha = 0.8$) and SDM ($\beta = 0.8$ and $\hat{M} = 4$) compared with the conventional method are illustrated in Fig 5.4. By using the HDM, the 2-nd, 5-th and 12-th segments are chosen from channels C2, C2, and CP4, respectively, while the rests are chosen from channel C4. By using the SDM, all children segments except the 2-nd, 5-th and 12-th segments are chosen from channel C4. The 2-nd segment is obtained by the weight average of channels C2, FC2, CP2 and FC4. The 5-th segment is obtained by the weight average of channels C2, CP2, C4 and FC4. The 12-th segment is obtained by the weight average of channels CP4, C2, CP2 and C4. After that, by using the merge and divide procedure, the resulting nonuniform segments obtained from using the HDM and SDM compared with the conventional method can be illustrated in Fig 5.5. The HDM and SDM result in visually clear ERD/ERS patterns in Fig 5.4. Consequently, more efficient nonuniform segments so that the ERD/ERS patterns are clearly partitioned can be observed in Fig 5.5. In addition, HDM and SDM yield similar nonuniform segments.

5.5.2 Classification Accuracy

Table 5.1 illustrates the classification accuracies of the HDM and SDM methods compared with the conventional method in [3]. Using the SDM yields slightly higher average classification accuracy than the HDM. Furthermore, using both the SDM and HDM outperform the conventional method by an average of 3% and more than 5% in some subjects.

Moreover, by replacing the LCP in Fig. 5.2(a) with the local MVDR packets, classification accuracy of subject 9 is shown in Table 5.2. Since the local MVDR packets transform results in data dependent bandpass filters with respect to each frequency of interest, it is reasonable that using the local MVDR packets yields around 4% higher classification accuracy than using the LCP. It should be noted that, even though, using the local MVDR packets leads to the improvement in classification accuracy, the compu-

Table 5.1. Classification accuracy (Acc.) of the HDM and SDM methods compared with the method in [3] denoted as conventional (NoF denotes number of features).

	HDM		SDM		Conventional	
Subjects	Acc.(%)	NoF	Acc.(%)	NoF	Acc.(%)	NoF
S1	78.33	6	78.89	15	77.22	14
S2	90	14	90.55	15	89.44	14
S3	74.44	20	75.56	4	72.22	5
S4	72.78	11	71.11	16	67.22	17
S5	70	9	68.89	10	67.78	9
S6	82.22	8	81.67	8	78.89	9
S7	87.78	22	88.33	10	86.11	20
S8	68.89	16	70.56	8	65.00	13
S9	75.56	14	75	15	70.56	12
Average	77.78	13.33	77.84	11.22	74.94	12.56

Table 5.2. Classification accuracy (%) of subject 9 using the LCP and local MVDR packets.

NoF	10	12	14	16	18	20
LCP	67.78	70.56	70	69.44	68.89	68.89
MVDR	73.89	72.78	72.78	75	72.78	72.22

tational loads are quite high compared with the LCP. Reducing the complexity requires further investigations.

5.6 Summary

We have presented a data dependent feature extraction scheme for classification of a left/right imagery multi-channel EEG called MF-LDB. Two methods, called HDM and SDM, are proposed for designing the MF-LDB. Besides using two fixed channels, these methods also employ other neighboring channels resulting in improvements of classification accuracies over the conventional scheme. The improvements from the proposed scheme support the previous studies that the ERD/ERS patterns may not occur at the

same position in different frequency bands. Furthermore, we have presented a novel local MVDR packets transform which is designed based on input data rendering highly selective frequency responses. Since the use of the local MVDR packets instead of the LCP leads to the improvement in classification accuracy, frequency band selection has a direct effect on extracting the important features of the ERD/ERS patterns.

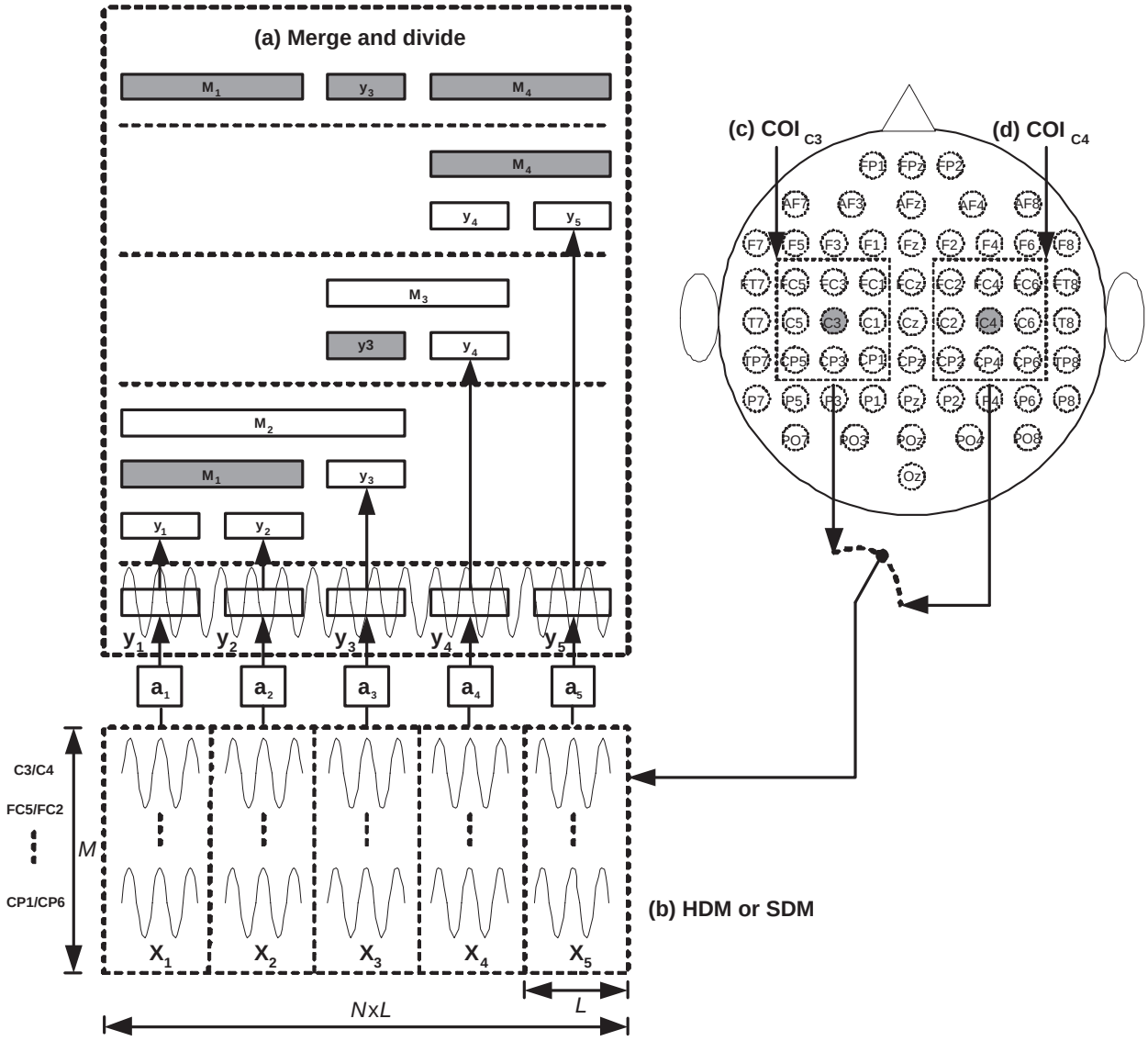

Figure 5.3. Design process of the MF-LDB: (a) merge and divide procedure, (b) HDM or SDM, (c) COI_{C3}, and (d) COI_{C4}.

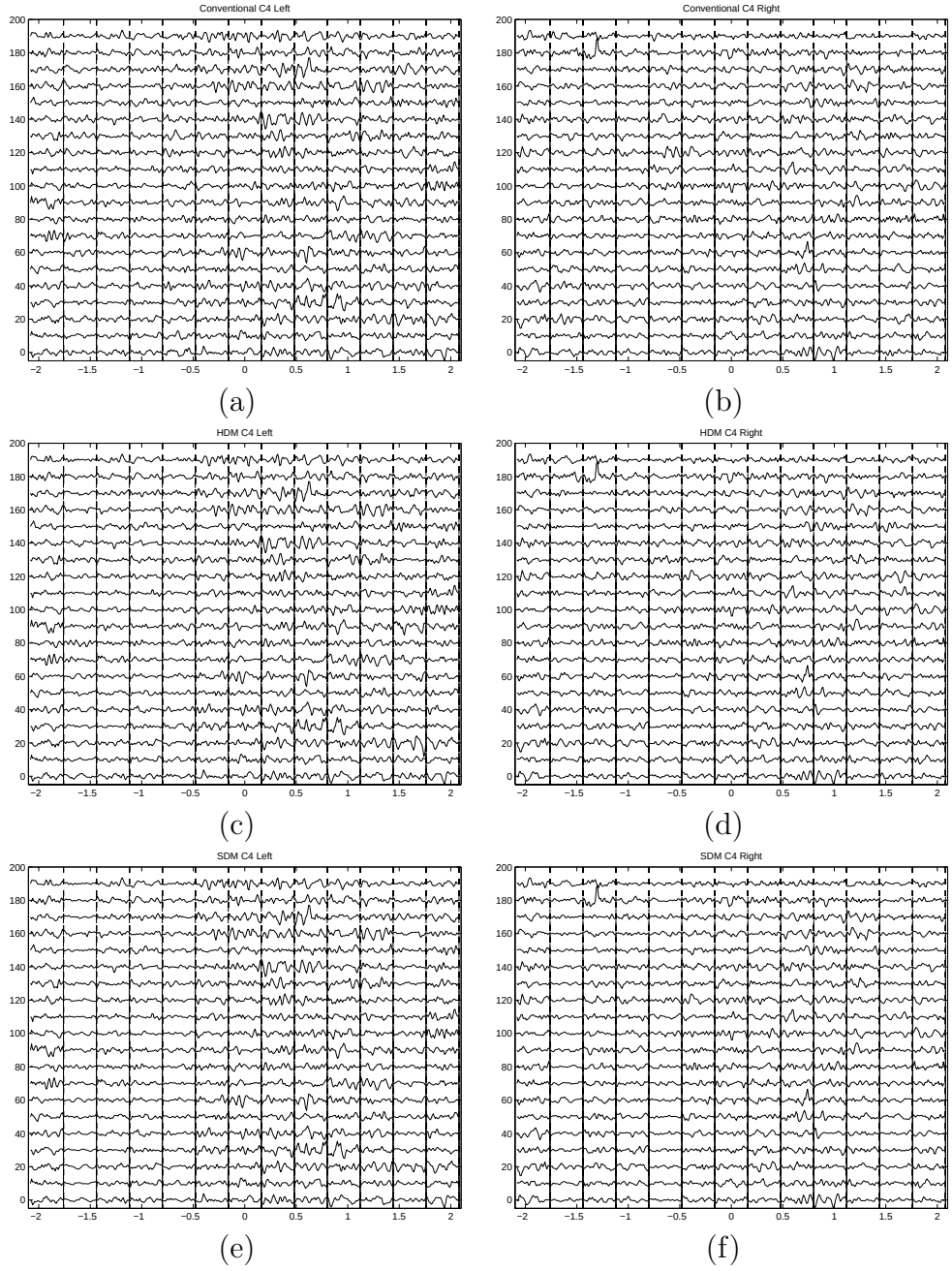

Figure 5.4. Original children segments of channel C4 of (a) left, (b) right imagery signals, resulting children segments using HDM of channel C4 and its neighboring channels of the (c) left, (d) right imagery signals, resulting children segments using SDM of channel C4 and its neighboring channels of the (e) left, (f) right imagery signals. Only 20 trials (with the mean shift by a multiple of 10) of each type of signals are shown for better visualization. x-axis represents the 4.16-second time interval as shown in Fig. 5.1, where 0 corresponds to 3.83 second in Fig. 5.1.

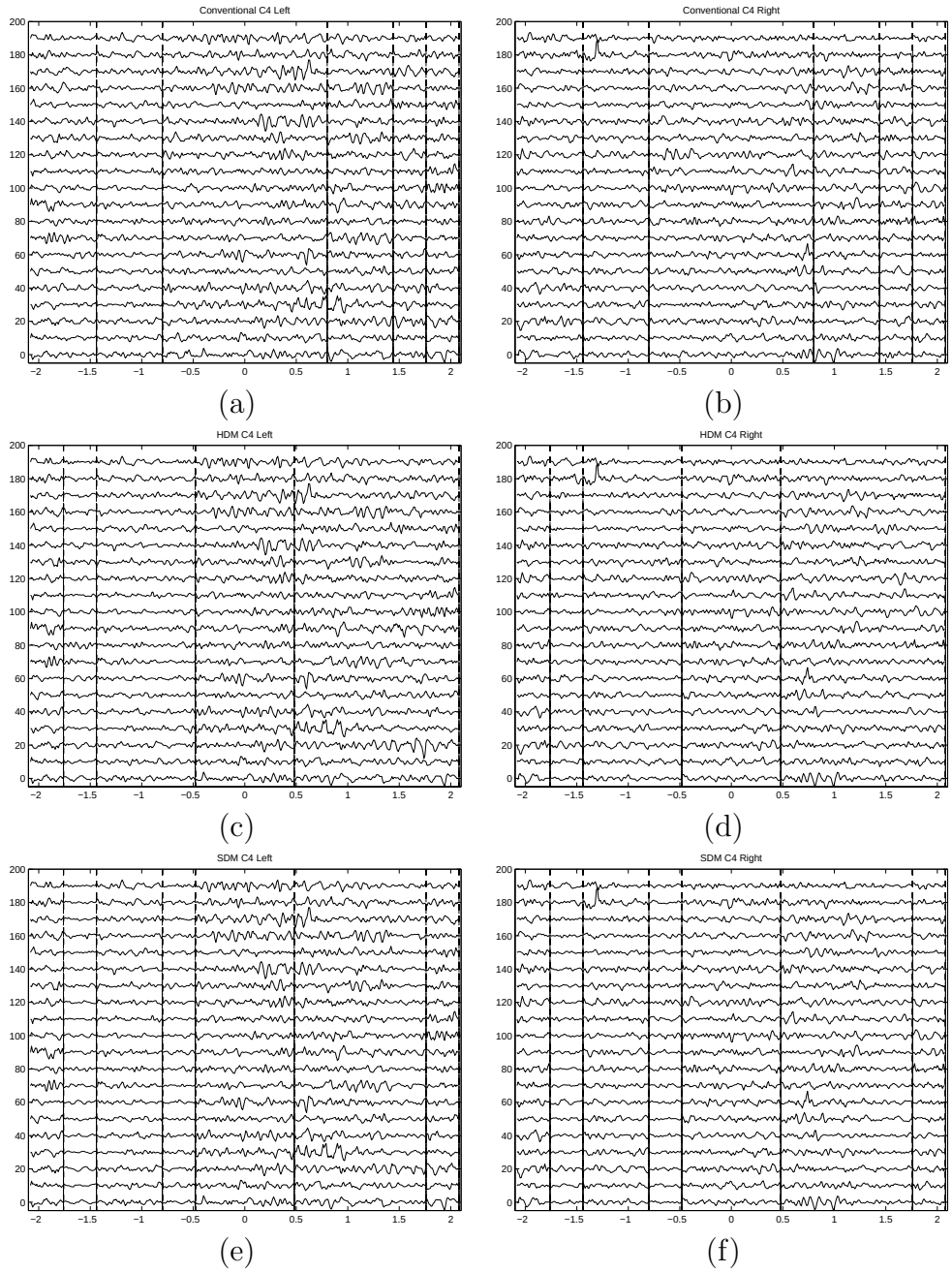

Figure 5.5. Resulting nonuniform segmentation (used for constructing the MF-LDB) after applying the merge and divide procedure to the signals in (a) Fig.5.4(a), (b) Fig.5.4(b), (c) Fig.5.4(c), (d) Fig.5.4(d), (e) Fig.5.4(e) and (f) Fig.5.4(f). Only 20 trials (with the mean shift by a multiple of 10) of each type of signals are shown for better visualization.x-axis represents the 4.16-second time interval as shown in Fig. 5.1, where 0 corresponds to 3.83 second in Fig. 5.1.

CHAPTER 6

CONCLUSIONS AND FUTURE DIRECTIONS

6.1 Conclusions

In this thesis, we have presented three investigations on signal processing for a multi-channel EEG. In the first investigation, we have presented a sub-optimal transform and used it for reducing the inter-channel correlation of the multi-channel EEG in the proposed lossless coder. The proposed transform is constructed by iteratively applying the KLT to the subdivided signals. Each small-sized KLT is then parameterized by lifting factorization resulting in a sub-optimal reversible transform called IntSKLT. In the second investigation, we have shown that a conventional STF model calculated from a 3-way PARAFAC can be approximated by its segmented versions, i.e. STF-TS, STF-SS, and STF-SS-TS models. We first segment the time domain, space domain, or both of them. After that the 4- or 5-way PARAFAC models is further applied. The signatures of the STF model can be well estimated by employing the resulting signatures from the reduced complexity models and the proposed formulae. The proposed models have been evaluated by employing in applications on artifact removal and left/right imagery EEG classification. In the last investigation, we have proposed a feature extraction scheme called MF-LDB which takes into account multiple-channel information of the EEG. The MF-LDB is obtained by calculating the LCP over the decided nonuniform segments. The decided segments are chosen from channels C3, C4, and their neighboring channels by the hard and soft decision. To further improve the classification results on the left/right imagery EEG classification, a non-redundant time-frequency transform which combines the concept of the minimum variance spectral estimation and the LCP have

been introduced. The proposed transform can significantly improve the classification results.

6.2 Future Directions

6.2.1 Seizure Prediction and Detection

Since structures of EEG signals depend highly on the situations and patients, the prediction and detection of some specific patterns simultaneously need both fast and sophisticated methods, especially for an online application. The success in this research can secure the many patients' life. Mostly, in this research area, time and frequency informations are usually employed. There are some other informations needed to be exploited, e.g. inter-channel, local channel (region), local time and phase. These informations can be put together by many methods. Optimal libraries can be designed for finding the patterns of each EEG channel. Moreover, the optimal parameters of wavelets or filters can be derived to obtain the optimal bases for representing each pattern of features. Mathematical formulae for the probabilistic models of the patterns of interest can also be found to improve the prediction and detection performances.

6.2.2 EEG Source Localization

Since EEG is the brain signal measured by a non-invasive method, it can be easily contaminated by the noises, artifacts (muscular, eye blink, speaking), and potentials from other positions in human body. Finding the true brain signals which represents the specific behaviors is an open question to the researchers. Since PARAFAC can be used for decomposing the signal into distinct components by taking into account the selected information, e.g. spatial, time, local time, and frequency. The PARAFAC model and the proposed reduced complexity version can be employed to solve this problem.

6.2.3 Reduced Complexity Space-Time-Frequency Model

Based on Chapter 4, the upper and lower bounds performances of the proposed reduced complexity space-time-frequency models need more investigation. Since the PARAFAC can be considered as one type of decomposition methods. The possible approach on finding the upper and lower bounds performances is to modify the model evaluation of the well known decompositions, e.g. independent component analysis (ICA), sparse component analysis (SCA), and principal component analysis (PCA). Moreover, an investigation on using the proposed reduced complexity models for the regression problem is also listed as the future works.

6.2.4 Emotion Recognition

To satisfy the need of human, designing a robot is a very interesting issue. In order to obtain the most efficient design, recognizing the emotion is needed. Emotion can be expressed by many ways, e.g. faces, sounds, heart rates, blood pressures, and brain signals (EEG). Each type of signals can be used to construct the dictionaries indicated its optimal patterns. There are many frameworks which can be modified based on this type of applications, e.g. matching pursuit, basis pursuit, method of frames, and best basis algorithm. Furthermore, the problem on combining more than one type of emotional signals and finding the optimal decision making also need to be formulated. Since some signals might not be considered as the useful information and can be considered as noises, we also need to find the optimal de-noising methods by taking statistical knowledge of each emotion into account.

APPENDIX A

ABBREVIATION LIST

1-D	One-dimensional
2-D	Two-dimensional
3-D	Three-dimensional
4-D	Four-dimensional
5-D	Five-dimensional
AAR	Adaptive autoregressive
ALS	Alternate least square
AR	Autoregressive
BCI	Brain computer interface
BP	Band power
BSS	Blind source separation
COI	Channel of interest
CORCONDIA	Core consistency diagnostic
DCT	Discrete cosine transform
DFT	Discrete Fourier transform
DLVQ	Distinction-sensitive learning vector quantization
DPCM	Differential pulse code modulation
DST	Discrete sine transform
EB	Eyeblink
ECG	Electrocardiogram
EEG	Electroencephalogram
EOG	Electrooculogram
ERD/ERS	Event related desynchronization/Event related synchronization
F-LDB	Flexible local discriminant bases
HDM	Hard decision making method

ICA	Independent component analysis
IntDCT	Integer discrete cosine transform
IntKLT	Integer Karhunen-Loeve transform
IntSKLT	Integer sub-optimal Karhunen-Loeve transform
KLT	Karhunen-Loeve transform
L/R	Left/right
LCP	Local cosine packets
LDA	Linear discriminant analysis
LDB	Local discriminant bases
LVQ	Learning vector quantization
MF-LDB	Multi-channel local discriminant bases
MVDR	Minimum variance distortionless response
PARAFAC	Parallel factor analysis
PCA	Principal component analysis
RMVB	Robust minimum variance beamformer
ROI	Region of interest
SDM	Soft decision making method
SERM	Single-row elementary reversible matrix
STF	Space-time-frequency
STF-SS	Space-time-frequency-space/segment
STF-TS	Space-time-frequency-time/segment
STF-SS-TS	Space-time-frequency-space/segment-time/segment
SVD	Singular value decomposition
SVM	Support vector machine
TERM	Triangular elementary reversible matrix
VQ	Vector quantization

REFERENCES

[1] S. Sanei and J. A. Chambers, *EEG Signal Processing*. West Sussex, England: John Wiley and Sons, 2007.

[2] L. Galli and S. Salzo, "Lossless hyperspectral compression using KLT," in *Proceedings of IEEE International Geoscience and Remote Sensing Symposium*, Sept. 2004, pp. 313–316.

[3] N. F. Ince, S. Arica, and A. H. Tewfik, "Classification of single trial motor imagery EEG recordings with subject adapted non-dyadic arbitrary time-frequency tilings," *Journal of Neural Engineering*, vol. 3, pp. 235–244, Sep. 2006.

[4] E. Niedermeyer and F. L. D. Silva, *Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*. Philadelphia, PA: Lippincott Williams and Wilkins, 2005.

[5] K. R. Rao and P. C. Yip, *The Transform and Data Compression Handbook*. Boca Raton, FL: CRC Press, 2001.

[6] P. Hao and Q. Shi, "Matrix factorizations for reversible integer mapping," *IEEE Transactions on Signal Processing*, vol. 49, pp. 2314–2324, Oct. 2001.

[7] R. Bro, *Multi-way Analysis in the Food Industry: Models, Algorithms and Applications*. MATLAB toolbox available [online] at http://www.models.kvl.dk/users/rasmus/: Ph.D. Thesis, University of Amsterdam and Royal Veterinary and Agricultural University.

[8] L. Cohen, *Time-Frequency Analysis*. Upper Saddle River, NJ: Prentice Hall, 1995.

[9] *Time-Frequency Toolbox*, Available [Online] at http://tftb.nongnu.org/.

[10] F. Miwakeichi, E. Martinez-Montes, P. A. Valdes-Sosa, N. Nishiyama, H. Mizuhara, and Y. Yamaguchi, "Decomposing EEG data into space-time-frequency components using parallel factor analysis," *NeuroImage*, vol. 22, pp. 1035–1045, 2004.

[11] S. Mallat, *Wavelet Tour of Signal Processing*. New York: Academic Press, 1998.

[12] J. Benesty, J. Chen, and Y. Huang, "A generalized MVDR spectrum," *IEEE Signal Processing Letters*, vol. 12, pp. 827–830, Dec. 2005.

[13] P. Stoica and R. Moses, *Introduction to Spectral Analysis*. New Jersy: Prentice Hall, 2007.

[14] G. Antoniol and P. Tonella, "EEG data compression techniques," *IEEE Transactions on Biomedical Engineering*, vol. 44, pp. 105–114, Feb. 1997.

[15] N. Memon, X. Kong, and J. Cinkler, "Context-based lossless and near-lossless compression of EEG signals," *IEEE Transactions on Information Technology in Biomedicine*, vol. 3, pp. 231–238, Sept. 1999.

[16] K. R. Rao and J. J. Hwang, *Techniques and Standards for Image, Video, and Audio Coding*. Upper Saddle River, NJ: Prentice Hall, 1996.

[17] *GZIP Software*, Available [Online] at http://www.gzip.org.

[18] P. Augustyniak, "Transform-based reduction of inter-channel correlation applied to lossless coding of multichannel ECG," in *Proceedings of International Conference of the IEEE Engineering in Medicine and Biology Society*, Oct. 2001, pp. 1966–1969.

[19] H. Hotelling, *Analysis of a Complex of Statistical Variables into Principal Components*. Baltimore: Warwick and York, 1933.

[20] K. I. Diamantaras and S. Y. Kung, *Principal Component Neural Networks: Theory and Applications*. New York: John Wiley and Sons, 1996.

[21] G. Strang and T. Q. Nguyen, *Wavelets and Filter Banks*. Cambridge, MA: Wellesley, 1996.

[22] M. Loeve, *Probability Theory*. Princeton: Van Nostrand, 1963.

[23] A. D. Poularikas, *The Transforms and Applications Handbook*. Boca Raton, FL: CRC Press, 2000.

[24] D. Yang, H. Ai, C. Kyriakakis, and C.-C. J. Kuo, "High-fidelity multichannel audio coding with Karhunen-Loeve transform," *IEEE Transactions on Speech and Audio Processing*, vol. 11, pp. 365–380, July 2003.

[25] P. Hao and Q. Shi, "Reversible integer KLT for progressive-to-lossless compression of multiple component images," *Proc. IEEE Int. Conf. Image Processing*, vol. 1, pp. 633–636, Sept. 2003.

[26] M. Gastpar, P.-L. Dragotti, and M. Vetterli, "The distributed Karhunen-Loeve transforms," *IEEE Transactions on Information Theory*, pp. 5177–5196, Dec. 2006.

[27] Y. Yokotani, R. Geiger, G. Schuller, S. Oraintara, and K. R. Rao, "Lossless audio coding using the IntMDCT and rounding noise shaping," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 14, pp. 2201–2211, Nov. 2006.

[28] Y. Wongsawat, S. Oraintara, T. Tanaka, and K. R. Rao, "Lossless multi-channel EEG compression," in *Proceedings of IEEE International Symposium on Circuits and Systems*, May 2006, pp. 1611–1614.

[29] Y. Wongsawat, S. Oraintara, and K. R. Rao, "Sub-optimal Integer Karhunen-Loeve transform for multichannel lossless EEG compression," in *Proceedings of European Signal Processing Conference*, Sept. 2006.

[30] Y. Wongsawat and S. Oraintara, "Sub-optimal KLT and its reversible approximation for lossless multi-channel EEG compression," *Submitted to Elsevier Transactions on Biomedical Signal Processing and Control*.

[31] P. P. Vaidyanathan, *Multirate Systems and Filterbanks*. Englewood cliffs, NJ: Prentice Hall, 1993.

[32] Y. J. Chen, S. Oraintara, T. Tran, K. Amaratunga, and T. Q. Nguyen, "Multiplierless approximation of transforms with adder constraint," *IEEE Signal Processing Letters*, vol. 9, pp. 344–347, Nov. 2002.

[33] *Shorten Software*, Available [Online] at http://www.etree.org/shncom.html.

[34] *JPEG2000 Software*, Available [Online] at http://www.kakadusoftware.com.

[35] G. Pfurtscheller and C. Neuper, "Motor imagery and direct brain-computer communication," *Proceeding of the IEEE*, vol. 89, pp. 1123–1134, Jul. 2001.

[36] L. D. Silva, "Analysis of eeg nonstationarities," *Electroencephalography and Clinical Neurophysiology*, vol. 34, pp. 163–179, 1978.

[37] S. Makeig, "Auditory event-related dynamics of the EEG spectrum and effects of exposure to tones," *Electroencephalography and Clinical Neurophysiology*, vol. 86, pp. 283–293, 1993.

[38] O. Bertrand, J. Bohorquez, and J. Pernier, "Time-frequency digital filtering based on an invertible wavelet transform: an application to evoked potential," *IEEE Transactions on Biomedical Engineering*, vol. 41, pp. 77–88, Jan. 1994.

[39] T. D. Lagerlund, F. W. Sharbrough, and N. E. Busacker, "Spatial filtering of multichannel electroencephalographic recording through principal component analysis by singular value decomposition," *Journal of Clinical Neurophysiology*, vol. 14, pp. 73–82, Jan. 1997.

[40] A. C. Soong and P. Z. Koles, "Principal-component localization of sources of the background EEG," *IEEE Transactions on Biomedical Engineering*, vol. 42, pp. 59–67, Jan. 1995.

[41] A. Cichocki and S. Amari, *Adaptive Blind Signal and Image Processing*. New York: John Wiley and Sons, 2002.

[42] A. Hyvarinen, J. Karhunen, and E. Oja, *Independent Component Analysis*. New York: John Wiley and Sons, 2001.

[43] A. Gonzalez, R. G. Menendez, C. M. Lantz, O. Blank, C. M. Michel, and T. Landis, "Non-stationary distributed source approximation: an alternative to improve localization procedures," *Human brain mapping*, vol. 14, pp. 81–95, 2001.

[44] T. Koenig, F. Marti-Lopez, and P. A. Valdes-Sosa, "Topographic time-frequency decomposition of EEG," *NeuroImage*, vol. 14, pp. 383–390, 2001.

[45] R. A. Harshman, "Foundations of the PARAFAC procedure: models and conditions for an explanatory multi-modal factor analysis," *UCLA Working Papers in Phonetics*, vol. 16, pp. 1–84, 1970.

[46] Y. Wongsawat, S. Oraintara, and K. R. Rao, "Reduced complexity space-time-frequency model for multi-channel EEG and its applications," in *Proceedings of IEEE International Symposium on Circuits and Systems*, May 2007, pp. 1305–1308.

[47] Y. Wongsawat and S. Oraintara, "Reduced complexity space-time-frequency model for multi-channel EEG," *Submitted to IEEE Transactions on Biomedical Circuits and Systems*.

[48] K. Nazarpour, Y. Wongsawat, S. Sanei, S. Oraintara, and J. Chambers, "A robust minimum variance beamforming approach for the removal of the eye-blink artifacts from EEGs," in *Proceedings of International Conference of the IEEE Engineering in Medicine and Biology Society*, May 2007.

[49] K. Nazarpour, Y. Wongsawat, S. Sanei, J. A. Chambers, and S. Oraintara, "Removal of the eye-blink artifacts from EEGs via robust minimum variance beamforming," *Submitted to IEEE Transactions on Biomedical Engineering*.

[50] A. K. Smilde, R. Bro, and P. Geladi, *Multi-Way Analysis with Applications in the Chemical Sciences*. West Sussex, England: John Wiley and Sons, 2004.

[51] M. Morup, L. K. Hansen, C. S. Herrmann, J. Parnas, and S. M. Arnfred, "Parallel factor analysis as an exploratory tool for wavelet transformed event-related EEG," *NeuroImage*, vol. 29, pp. 938–947, 2006.

[52] A. Osman and A. Robert, "Time-course of cortical activation during overt and imagined movements," in *Proceedings of Cognitive Neuroscience Annual Meeting*, 2001.

[53] K. Nazarpour, S. Sanei, L. Shoker, and J. Chambers, "Parallel space-time-frequency decomposition of EEG signals for brain computer interfacing," in *Proceedings of European Signal Processing Conference*, 2006.

[54] G. Pfurtscheller and F. H. L. da Silva, "Even-related EEG/MEG synchronization and desynchronization: basis principles," *Clinical Neurophysiology*, pp. 1842–1857, Nov. 1999.

[55] S. A. Vorobyov, A. B. Gershman, and Z. Q. Luo, "Robust adaptive beamforming using worst case performance optimization," *IEEE Transactions on Signal Processing*, vol. 51, no. 2, pp. 313–324, 2003.

[56] J. Li and P. Stoica, *Robust Adaptive Beamforming*. Hoboken, NJ: John Wiley and Sons, 2005.

[57] J. Kalcher, D. Flotzinger, and G. Pfurtscheller, "Brain-computer interface-a new communication device for handicapped persons," *Journal of Microcomputer Applications*, vol. 16, pp. 293–299, 1993.

[58] M. Prezenger and G. Pfurtscheller, "Frequency component selection for an EEG-based brain computer interface," *IEEE Transactions on Rehabilitation Engineering*, vol. 7, pp. 413–419, Dec. 1999.

[59] C. Anderson, E. Stolz, and S. Shamsunder, "Discriminating mental tasks using EEG represented by AR models," in *Proceedings of International Conference of the IEEE Engineering in Medicine and Biology Society*, 1995.

[60] D. J. McFarland, A. T. Lefkowicz, and J. R. Wolpaw, "Design and operation of an EEG-based brain-computer interface with digital signal processing technology," *Behavior Research Methods, Instruments, and Computers*, vol. 3, pp. 337–345, 1997.

[61] G. Pfurtscheller, C. Neuper, A. Schlogl, and K. Lugger, "Separability of EEG signals recorded during right and left motor imagery using adaptive autoregressive parameters," *IEEE Transactions on Rehabilitation Engineering*, vol. 6, pp. 316–325, Sep. 1998.

[62] N. F. Ince, A. H. Tewfik, and S. Arica, "Extraction subject-specific motor imagery time-frequency patterns for single trial EEG classification," *Computers in Biology and Medicine*, vol. 37, pp. 499–508, Apr. 2007.

[63] T. Wang, J. Deng, and B. He, "Classifying EEG-based motor imagery tasks by means of time-frequency synthesized spatial patterns," *Clinical Neurophysiology*, vol. 115, pp. 2744–2753, Dec. 2004.

[64] N. Saito, R. R. Coifman, F. B. Geshwind, and F. Warner, "Discriminant feature extraction using empirical probability density and a local basis library," *Pattern Recognition*, vol. 35, pp. 1842–1852, Dec. 2002.

[65] R. Duda, P. Hart, and D. Stork, *Pattern Classification.* New York: John Wiley and Sons, 2002.

[66] S. C. Chan, Z. G. Zhang, and K. M. Tsui, "Minimum variance spectral estimation-based time frequency analysis for nonstationary time-series," in *Proceedings of IEEE International Symposium on Circuits and Systems*, May 2007, pp. 1815–1818.

[67] Y. Wongsawat and S. Oraintara, "Multi-channel flexible local discriminant bases for left/right imagery EEG classification," *Submitted to Proceedings of IEEE International Symposium on Circuits and Systems*, May 2008.

[68] ——, "Left/right imagery EEG classification using multi-channel flexible local discriminant bases," *Submitted to IEEE Transactions on Neural Systems and Rehabilitation Engineering.*

[69] B. Hjorth, "An online transformation of eeg scalp potentials into orthogonal source derivations," *Electroencephalography and Clinical Neurophysiology*, vol. 39, pp. 526–530, Nov. 1975.

[70] S. Theodoridis and K. Koutroumbas, *Pattern Recognition*. San Diego: Academic Press, 1999.

BIOGRAPHICAL STATEMENT

Yodchanan Wongsawat was born in Chiangmai, Thailand, in 1979. He received his B.E. degree from Sirindhorn International Institute of Technology (SIIT), Thammasat University, Thailand, in 2001, his M.S. degree from the University of Texas at Arlington in 2003, both in Electrical Engineering. His research interests are in the area of biomedical signal processing, pattern recognition, time-frequency analysis, and data compression. He is a member of the Institute of Electrical and Electronics Engineers (IEEE) and European Association for Signal Processing (EURASIP) societies.