TENSOR FACTORIZATION FOR ENGINEERING APPLICATIONS

Researcher:

Muhammad Akmal

REG NO. 100-FET/PHDEE/F15

Supervisor:

Prof. Dr. Syed Zubair

Department of Electrical Engineering

Faculty of Engineering & Technology

INTERNATIONAL ISLAMIC UNIVERSITY,

ISLAMABAD

Tensor Factorization

Electromyography

Tensor factorization for Engineering Applications

Muhammad Akmal

100-FET/PHDEE/F15

Submitted in partial fulfillment of the requirements for the PhD degree in Electronic Engineering at the Department of Electrical Engineering

Faculty of Engineering and Technology

International Islamic University,

Islamabad

Supervisor

Prof. Dr. Syed Zubair

August, 2020

DEDICATED TO

My Teachers,

Parents,

And Wife.

CERTIFICATE OF APPROVAL

Title of Thesis: **Tensor factorization for Engineering Applications**

Name of Student: Muhammad Akmal

Registration No.: 100-FET/PhDEE/F15

This thesis is accepted by the Department of Electrical Engineering, Faculty of Engineering and Technology, International Islamic University, Islamabad, in partial fulfillment of the requirements for the PhD degree in Electronics Engineering.

VIVA VOCE COMMITTEE

Supervisor

Prof. Dr. Syed Zubair

Professor,

Department of Computer Science, University of Sialkot ____________________

Internal Examiner

Dr. Suheel Abdullah Malik

Associate Professor,

Department of Electrical Engineering, FET, IIUI ____________________

External Examiner 1

Prof. Dr. Ijaz Mansoor Qureshi

Professor,

DEE, Air University, Islamabad ____________________

External Examiner 2

Dr. Muhammad Mukhtar Talha

Principal Scientific Officer, KRL ____________________

Chairman DEE

Dr. Suheel Abdullah Malik

Associate Professor,

Department of Electrical Engineering, FET, IIUI ____________________

Dean FET

Prof. Dr. Muhammad Amir

Professor,

Department of Electrical Engineering, FET, IIUI ____________________

ABSTRACT

Tensor factorization in which a tensor (multidimensional array) is decomposed into a sum of rank-one tensors has been used extensively in various engineering applications. In this work, we apply tensor factorization to recover missing data in biomedical signals. Biomedical signals such as electromyography (EMG) are used to control robotic arms. During acquisition of biomedical signals, data is missed because of various reasons such as disconnection of electrodes, artifacts, muscle fatigue, or incapability of instruments to collect very low amplitude signals. Missing data recovery is an important application to maintain signals fidelity or otherwise classification accuracy of algorithms degrades.

Traditional algorithms for tensor factorization are canonical polyadic decomposition (CPD) and Tucker decomposition. The present work is an advancement in CPD to recover missing data with significantly improved accuracy in terms of relative mean error (RME). CPD decomposes the Nth-order tensor into a linear combination of N rank-1 tensors. The concept of CPD is exploited to develop CP based weighted optimization (WOPT) method in which only known data is modeled. RME using WOPT is reduced significantly as compared to other tensor factorization methods which shows good recovery of missing data. In this work, we compared performance of CP-WOPT with Non-negative Matrix Factorization (NMF) and PARAFAC/CPD. Performance of traditional methods such as NMF and CPD degrade as amount of missing data increase but CP-WOPT outperforms traditional methods even in worst case of 95% of unstructured (individual samples of data is missing randomly) and 50% of structured missing data (chunks of data is missing randomly). Our proposed framework has been tested on synthetic data, surface EMG and

intramuscular EMG data of both healthy and amputee subjects with two types of missing data schemes of unstructured and structured. For unstructured missing data, we have removed 60% to 95% data from total EMG data whereas for structured missing data we have removed 10% to 50% data from first and second halves for all four movements. Results show CP-WOPT outperformed both NMF and CPD to recover unstructured and structured missing data from synthetic, sEMG and iEMG data. Furthermore, in order to test the robustness of our framework we have employed CP-WOPT on large multiday iEMG data of healthy as well as amputee subjects. Results show that CP-WOPT even has outperformed NMF and CPD in multiday EMG data as well. The effective application of proposed framework on synthetic as well as real-life Electromyography (EMG) signals (which include both surface EMG and intramuscular EMG signals) establishes the worth and efficacy of the method. Prosthetic hand is one of the application where EMG data is collected to control it. Because of various issues, part of EMG data is lost which leads to poor classification of finger movements. In this work, missing data is recovered so that functioning of prosthetic hand could be improved. We do so by firstly acquiring EMG data from healthy and amputee subjects. Then this research builds on recovering structured and unstructured missing data by employing matrix and tensor factorization methods.

LIST OF PUBLICATIONS AND SUBMISSIONS

[1]. **M. Akmal,** S. Zubair, M. Jochumsen, E. N. Kamavuako and I. K. Niazi, "A tensor-based method for completion of missing electromyography data," IEEE Access, 2019. DOI: 10.1109/ACCESS.2019.2931371 **(IF 4.098)**

SUBMITTED PAPER

[1]. **M. Akmal,** S. Zubair, M. Jochumsen, M. Zia ur rehman, E. N. Kamavuako and I. K. Niazi, "Multiday evaluation of tensor factorization to recover missing data in intramuscular electromyography data." IEEE Transactions on biomedical engineering 2020.

The research work presented in this dissertation is based on the accepted publication 1 and submitted publication 1.

ACKNOWLEDGEMENTS

In the name of Allah (Subhanahu Wa Ta'ala), who is the most gracious and the merciful. I would like to thank Allah for giving me strength and patience to complete this research work. Peace and blessings of Allah be upon His last Prophet Muhammad (Sallulah-o-Alaihihe-Wassalam) and all his Sahaba (Razi-Allah-o-Anhu) who dedicated their lives for Dawah and spread of Knowledge.

I am truly grateful to my supervisor Dr. Syed Zubair whose inspiration, ideas and efforts make it possible for me to complete my higher studies. He has been a role model for me and many others in teaching, research and other aspects of life. I would also like to thank Dr. Imran Khan Niazi for his time, support and guidance during my research work.

I am thankful to administration at department, as well as university level for their kind support. I am really grateful to my father and mother sister for thei love and support throughout my life. I am also very thankful to my wife for her patience, encouragement and prayers during every stage of my PhD degree.

(Muhammad Akmal)

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATIONS

NMF	Nonnegative Matrix Factorization
PARAFAC	Parallel Factor
CPD	Candecomp / PARAFAC Decomposition
CP-WOPT	Canonical / Polyadic Weighted OPTimization
EEG	Electroencephalography
EMG	Electromyography
ECG	Electrocardiography
ALS	Alternating Least Squares
sEMG	Surface Electromyography
iEMG	Intramuscular Electromyography
RME	Relative Mean Error
PCA	Principal Component Anlaysis
FA	Factor Analysis
ICA	Independent Component Analysis
EMD	Empirical Mode Decomposition
LLGMN	Log Linearized Gaussian Mixture Network
NN	Neural Networks
SVM	Support Vector Machine

LIST OF SYMBOLS

A list of commonly used symbols in this dissertation are given below.

$\mathcal{X}$	Data tensor
$\mathcal{W}$	Weighting tensor
$\mathcal{N}$	Noise tensor
t	Time
$\mathbf{A}$	Matrix
$\mathbf{a}$	Vector

Chapter 1.

Introduction

1.1 Background and motivation

Electromyography (EMG) signal is the primary biomedical signal that is used for myoelectric control. It has two types 1) surface EMG (sEMG) and 2) intramuscular EMG (iEMG). Surface EMG is the signal captured from above the skin, whereas intramuscular EMG is the signal captured from inside skin. Sensors or electrodes are used to capture sEMG signals, whereas needles are used to capture iEMG signals. Acquisition of EMG signals need preprocessing and conditioning of data. There exists various control schemes to efficiently translate and utilize EMG data such as ON-OFF control, proportional control, finite state machine control, direct control etc. [1].

Use of prosthetic devices such as prosthetic hands, prosthetic limb etc. have increased drastically ever since advancements have been made in myoelectric devices and state-of-the-art artificial intelligence algorithms are employed in them. Prosthetic devices are controlled by EMG signals or more precisely by the electrical activity recorded from muscles. Artificial intelligence algorithms perform efficiently if input data (EMG in our case) is large in amount. However, it has been observed that EMG data is not lossless because of factors such as artifacts or disconnection of electrodes. Performance of prosthetic hand rely on acquired EMG data and artificial intelligence algorithms, whereas incomplete or missing EMG data reduces performance of artificial intelligence algorithms and consequently control of prosthetic hand. Therefore, if missing data is recovered efficiently than control of prosthetic hand can be improved significantly.

1.2 Research problem statement

It has been observed that during data collection, some of it is missed because of different reasons. For example, in EEG or EMG signals data is missed because of disconnection of electrodes, muscle fatigue, or incapability of instruments to collect very low amplitude data. Such incomplete data degrades the performance of myoelectric devices due to inaccurate execution of movements. In different fields of engineering, data is collected for various applications. For example, Electroencephalography (EEG), Electromyography (EMG) are collected in the field of biomedical engineering for the control of prosthetic arms or wheel-chairs, and communication for people with various motor disabilities. Artificial intelligence algorithms, which are employed in biomedical applications such as classification of movements in prosthetic arms or wheelchairs require accurate as well as complete data for optimal performance.

Standard practice is to fill missing values with different estimates, such as mean value. However, it has been observed that performance of such estimates degrades with the increase in missing data. Another method to recover missing data is based on matrix factorization in which multidimensional data is first flattened or unfolded and then factorized, which destroys intrinsic multi-way nature of data. Usually methods such as Alternating Least Squares (ALS), are used to minimize the objective function on matrix factorization. Therefore, in order to preserve multi-way nature of data, tensor factorization is employed to recover missing data. However conventional methods for tensor factorization also exhibit poor performance in estimating missing values. This is due to their modeling technique in which all known and missing values are used to estimate missing values [2]. In order to improve performance of recovering missing data, our study aims to model only known entries by formulating modelling technique as a weighted least squares problem. It is done by minimizing weighted error function $\frac{1}{2}\|\{(\mathcal{X} - [\![\mathbf{A}^{(1)}\mathbf{A}^{(2)}, \ldots, \mathbf{A}^{(N)}]\!])\mathcal{W}\}\|_{\mathcal{X}}^{2}$ between original $\mathcal{X}$ and recovered data $\overline{\mathcal{X}}$. Research problem parameters include $\mathcal{Z}$ (which has weighted tensor multiplied with recovered tensor), f_w (computed function), and $G^{(n)}$ (gradients). Moreover, the proposed framework is tested on both surface and intramuscular electromyography (EMG) data of healthy and amputee subjects. Therefore, tensor factorization is employed effectively for engineering application to fill the EMG data so that performance of prosthetic hand can be improved.

1.3 Research objectives

The following research work has two major objectives:

- Design of novel tensor factorization framework based on strong mathematical foundations of multilinear algebra.

- Applications of proposed tensor factorization framework to real-life surface and intramuscular EMG data of both healthy and amputee subjects.
- To apply the proposed tensor factorization framework to EMG data for preferential accuracy of the proposed technique.

1.4 Research philosophy

Missing data in EMG signals degrades control performance of prosthetic devices and reduces accuracy of model or leads to biased model. To improve control of prosthetic hand or limb, missing data should be recovered efficiently. To recover missing data, tensor and matrix factorization methods are employed. Both tensor and matrix factorization methods discover latent features, which are then used to predict missing values and thereby improve the performance of artificial intelligence algorithms. Hence, in this thesis, we will investigate application of tensor factorization on biomedical signals to recover missing values for better reconstruction of EMG signals.

1.5 Hypothesis

Movement of prosthetic hand is controlled by classification algorithms such as linear discriminant analysis (LDA), support vector machine (SVM) etc. Performance of classification methods rely on amount of data i.e. classification accuracy improves if amount of input data is large. Whereas real-life EMG signals are not lossless and considerable amount of data is missed, which degrades accuracy of classification methods. To improve accuracy of classification methods, amount of input data should be increased by employing different methods such as matrix and tensor factorization methods. Performance of matrix and tensor factorization methods is measured by performance metric i.e. relative mean error (RME). For this research, ten subjects (all male) were recruited for EMG data acquisition. Ages of all subjects ranged from 18 to 38 years old (mean ± standard deviation (SD), 24.5±2.3y). All subjects were healthy with no neuromuscular disorders.

1.6 Research methodology

The methods employed to recover missing data on simulated and real-life EMG data are based on matrix and tensor factorization. The objective function of both matrix and tensor factorization methods is to find the hidden factor matrices of the incomplete input data. In matrix factorization method, input data is matricized so that it is applicable to it. Whereas, in tensor factorization methods input data is kept in its original multi-dimensional form. Traditionally matrix and tensor factorization methods model the entire input data including missing values and thus a weak model is built. Such weak model does not recover missing data efficiently. Therefore, our proposed framework models only the known entries of input data. Where model is formulated as weighted least squares problem and solved by a gradient descent optimization approach.

To test the ability of proposed framework to recover missing data, it was employed on simulated data of various sizes, surface EMG (sEMG) data of healthy subjects, and intramuscular EMG (iEMG) data of both healthy and amputee subjects. Proposed framework outperformed traditional state-of-the-art methods on single day sEMG data. To test the performance of proposed framework on large data, it was employed on multiday iEMG data. Our proposed framework significantly outperformed traditional state-of-the-art methods in large multiday data as well.

Research objective	Remarks
Design of novel tensor factorization framework based on strong	Research objective achieved

mathematical foundations of multilinear algebra.	
Applications of proposed tensor factorization framework to real-life surface and intramuscular EMG data of both healthy and amputee subjects.	Research objective achieved
To apply the proposed tensor factorization framework to EMG data for preferential accuracy of the proposed technique.	Research objective achieved

1.7 Thesis outlines

This thesis is organized as consisting of five chapters. In Chapter 1 is presented a conceptual outline of the whole thesis by mentioning a background to identify Research Problem. The research problem statement with objectives and significance. The philosophy and hypothesis make the two pillars upon which the framework is built. Literature review is given in Chapter 2 which presents the current research relevant to this thesis. Methodology is discussed in Chapter 3 which presents the methodological framework employed to achieve the research objectives. Results of proposed framework are discussed in Chapter 4, where single day surface EMG data of healthy subjects and multiday intramuscular EMG data of healthy as well amputee subjects was utilized. Conclusions and further work is provided in Chapter 5.

Chapter 2.

Literature Review

2.1 Introduction

This study discusses the recovery of missing data in sEMG signals that arise during the acquisition process. Missing values in EMG signals occur due to either disconnection of electrodes, artifacts, muscle fatigue or incapability of instruments to collect very low amplitude signals. In many real-world EMG related applications, algorithms need complete data to make accurate and correct predictions, or otherwise, the performance of prediction reduces sharply. Tensor factorization methods are employed to recover unstructured and structured missing data from EMG signals. In this work, we use first-order weighted optimization (WOPT) of PARAFAC decomposition model to recover missing data.

2.2 General Background

ELECTROMYOGRAPHY (EMG) is a diagnostic technique which records the electrical activity produced by contraction of muscles. The electric activity or potential is generated by the muscle cells when these cells are electrically activated. Generally, two types of EMG exist surface EMG (sEMG) and intramuscular EMG (iEMG). sEMG is the recording of electrical activity from the muscle surface (non-invasive), whereas iEMG is recorded directly within the muscle tissue. EMG signals have many applications such as upper-limb prostheses [3], [1],[4] , electric wheelchairs control [5] and muscle-computer interaction [6]. In these applications, complete EMG signals without missing data are required for efficient and successful implementation. However, practically, EMG data acquisition is not lossless. During signal acquisition, data is lost due to many reasons, such as artifacts or disconnection of electrodes with the body [7]. These missing values in the EMG signal can cause degradation in the overall performance of healthcare applications, such as myoelectric pattern recognition to predict motor intention from sEMG signals [7]. Moreover, missing values also reduce the accuracy of the classification of movements for prostheses control [8]. If data is incomplete and the percentage of missing data is large, then the classification performance and statistical power of those classification methods highly degrade, which makes it important to have complete data set. To effectively estimate the missing data, proper imputation methods must be utilized. Generally, in EMG applications, missing data had either not been recovered or estimated by simply replacing it with mean values of the neighboring data values, which has proven to be highly sub-optimal [9]. In this study, we focus on estimating missing values using multidimensional data structure [2], [10] based upon multilinear algebra (tensors).

This study aims to recover missing values in surface EMG signals by estimating the latent structure of the data. In order to estimate latent structure, tensor factorization methods have been employed, which produce factor matrices which are used to produce the reconstructed tensor. A weighted version of an error function has been further formulated that ignores the missing values and model only the known values which improve the estimation accuracy of recovering missing data significantly.

Matrix and Tensor decomposition of EMG signals have been widely studied in the literature. In [11-15], non-negative matrix factorization (NMF) has been applied on EMG signals for various applications, e.g. recognition of gestures, to obtain information for neural control and identification of various surface EMG signals. In [16], various matrix factorization algorithms such as Principal Component Analysis (PCA), Factor Analysis (FA), Independent Component Analysis (ICA), and Non-negative Matrix Factorization (NMF) were evaluated on EMG recording. In [17], surface EMG signals are decomposed using non-negative Tensor factorization to find the features for classification purpose. In [18], NMF was employed to identify EMG finger movements to evaluate the functional status of hand so that it can assist in hand gesture recognition, prosthetics and rehabilitation applications. In [19], FastICA method is implemented for EMG signals decomposition. In [20], NMF along with different initialization techniques was applied to acquire muscle synergies which are important for generating biomechanical tasks. In [21], higher order tensor decompositions are employed on EMG signals to estimate muscle synergies.

Electromyography (EMG) has many daily-life applications e.g. EMG controlled prosthetic limb [22], EMG based embedded system [23] to control a six degree of freedoms (DOFs) prosthetic hand, [24] presents an extensive review on control strategies of prosthetic hands, whereas in [25] an EMG-based cost-effective design of prosthetic hand is proposed. EMG signals have two types 1) surface EMG signals and 2) Intramuscular EMG signals. Surface EMG (sEMG) signals are acquired from electrodes that are mounted on the skin whereas intramuscular EMG (iEMG) signals are acquired from needle electrodes inserted through the skin into muscle tissue. Although sEMG signals are widely used, recordings are highly variable because the innermost layer of skin is made of fat. In order to obtain better EMG signals, iEMG signals were explored, which are highly specific with respect to what muscle signals are recorded from [26]. Moreover, it allows the recording of EMG signals from the

Missing data degrades the accuracy of classification methods because they do not analyze the relationship with other variables correctly [52]. Usually, there are two ways to improve the performance of classification methods: 1) by increasing the total number of samples or input data because a large data set relies less on assumptions, 2) by recovering missing data efficiently by replacing zeroes with useful values. In [53], it is shown that filling the missing values improves performance of classifiers such as KNN, SVM, etc. In [54], four methods (case deletion, mean imputation, median imputation, and knn-imputation) are compared to impute missing data and their effects are shown on classification performance. They claim that all methods perform better after imputation.

2.3 Contemporary research closer to background

However, so far in the literature, missing data in EMG signals has been recovered by using ensemble classifier systems (Benchmark paper 1) [9], nonlinearities interpolation approach (Benchmark paper 2) [32], mean data imputing [7], Empirical Decomposition Mode (EMD) [55] and marginalization and conditional-mean imputation [31]. In [9], imputation and reduced-feature models were employed to perform classification in presence of missing data but the results were not promising. In [32], missing data of up to 80% was recovered. However they tested algorithm on single subject and it is also unclear whether they recovered unstructured or structured missing data. In [7], imputation was carried out using mean of data which works poorly on non-stationary EMG data. In [55], EMD fails to recover structured missing data. In [31], the main focus was on developing classification model. However they also employed a simple mean imputation method to recover missing data. In [2] and [56], tensor factorization techniques are applied on EEG signals. However, so far, EMG signals have not been explored that way. For the first time, in this work, missing data is recovered in EMG signals with a detailed analysis in which matrix, as well as tensor factorization methods, are employed. We apply NMF for matrix factorization and, PARAFAC and CANDECOMP/PARAFAC - Weighted OPTimization (CP-WOPT) for tensor factorization. As normalized EMG data contains non-negative values; hence, for the case of matrix factorization we apply NMF, which is the unsupervised learning algorithm used for dimensionality reduction and construction of low-dimensional approximation of observed data. NMF is more suitable because other methods such as Principal Component Analysis (PCA) produce the factors which can be positive or negative. To our knowledge, tensor factorization for recovering missing data in EMG signals has not been studied yet. In this work, for the first time, we employ the tensor factorization method to recover unstructured and structured missing data in EMG signals. We apply PARAFAC and weighted optimization (WOPT) of PARAFAC model to EMG signals and recover missing data efficiently as compared to matrix factorization techniques. In [33], the recovery of missing data in surface Electromyography (sEMG) signals is discussed that arise during the acquisition process. Missing values in EMG signals occur due to either disconnection of electrodes, artifacts, muscle fatigue or incapability of instruments to collect very low amplitude signals. In many real-world EMG related applications, algorithms need complete data to make accurate and correct predictions, or otherwise the performance of prediction reduces sharply. Matrix and tensor factorization methods are employed to recover unstructured and structured missing data from EMG signals. First order weighted optimization (WOPT) of PARAFAC decomposition model is used to recover missing data. Proposed framework is tested against Non-Negative Matrix Factorization (NMF) and Parallel Factor Analysis (PARAFAC) on simulated as well as on offline EMG signals having unstructured missing values (randomly missing data ranging from 60% to 95%) and structured missing values. In the case of structured missing data havi..g different channels, the percentage of missing data of a channel goes up to 50% for different movements. It has been observed empirically that the proposed framework recovers the missing data with relatively much improved accuracy in terms of Relative Mean Error (up to 50% and 30 % for unstructured and structured missing data respectively) as compared to matrix factorization methods even when the portion of unstructured and structured missing data reaches up to 95% and 50%, respectively. Missing data in surface EMG signals are recovered using matrix and tensor factorization-based methods. Moreover, their performances are compared to recover missing data in noisy simulated data and real-world EMG data to show that the tensor-based approach outperforms matrix factorization based approach. Problem of missing data in extreme cases is also addressed when up to half consecutive EMG samples of a particular channel are missing. Our proposed framework successfully recovers the missing data even in such an extreme case as well.

Missing data from intramuscular EMG signals of both healthy and amputee subjects are also recovered using state-of-the-art tensor factorization methods. Performance of matrix and tensor factorization methods is compared to recover missing data in real-world iEMG signals as well. Furthermore, multiday EMG data is utilized to test the performance of both matrix and tensor factorization methods in a large multiday iEMG dataset. Extreme case is also considered when up to 50% iEMG data is missing from day 1 to 7 and performance of both matrix and tensor factorization methods is tested. Results show that CP-WOPT outperformed both NMF and CPD to recover missing data even in the worst-case scenario in iEMG data as well.

2.4 Summary

This chapter has described the general background of the research, identifying clearly the importance of EMG signals in biomedical engineering. The use of EMG for applications in prosthetic parts of the body has been discussed. Techniques for EMG signals acquisition has been discussed, and sources of information from EMG signal data has been presented. This is done to explore how such missing elements of data could affect the overall performance of the prosthetic parts. Algorithms to this effect are critically reviewed in benchmark 1 and benchmark 2 for identifying the dark spots yet to be highlighted. A ground and reason for this research is established by continuing from benchmark 1 and benchmark 2 onward for formulating modelling technique as a weighted least squares problem as the core framework in this research, which is elaborated in Chapter 3 for achieving the target objectives.

Chapter 3.

Methodological Framework Details

3.1 Introduction

This chapter presents the material and methodology adopted for all two studies. The first study covers single day surface EMG signals of ten healthy patients. The second study covers multiday intramuscular EMG data of both healthy and amputee subjects.

3.2 Notations and preliminaries

Tensor $\mathcal{X}_{(i,j,k...)}$ is a multi-dimensional array which has different modes for data representation. A tensor with one mode is a one-dimensional array referred to as a vector and with two modes is known as the matrix. A tensor of third order is shown in Fig. 3.1, which has three dimensions having indices $i = 1, \dots, I$, $j = 1, \dots, J$ and $k = 1, \dots, K$. In the current work, a tensor is represented by uppercase Blackadder ITC letter $\mathcal{X}$, a matrix is represented by bold italic uppercase letter $\boldsymbol{X}$, a vector is denoted by italic bold lower case letter $\boldsymbol{x}$, and a scalar is represented by italic lowercase letter x. The individual elements of nth-order tensor are represented by lowercase letters with subscripts e.g. if N-way tensor has $(I_1 \times I_2 \times \dots \times I_N)$ samples then its *n-th* element is denoted by $x_{i_1 i_2 \dots i_N}$.

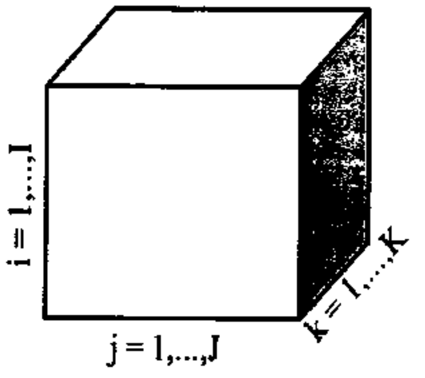

Fig. 3.1 Tensor of third order: $\mathcal{X} \in \mathbb{R}^{I \times J \times K}$

The *Scalar product* of two tensors $\mathcal{X}$, $\mathcal{Y}$ with size $I_1 \times I_2 \times \ldots \times I_N$ is defined as:

$$\langle \mathcal{X}, \mathcal{Y} \rangle = \sum_{i_1} \sum_{i_2} \cdots \sum_{i_N} x_{i_1 i_2 \ldots i_N} y_{i_1 i_2 \ldots i_N}$$

The *Hadamard product* of two tensors $\mathcal{X}$, $\mathcal{Y}$ is defined as:

$$(\mathcal{X} * \mathcal{Y})_{i_1 i_2 \ldots i_N} = x_{i_1 i_2 \ldots i_N} y_{i_1 i_2 \ldots i_N}$$

The *Frobenius norm* of a tensor $\mathcal{X}$ is given by:

$$\|\mathcal{X}\|_F = \sqrt{\sum_{i_1=1}^{I_1} \sum_{i_2=1}^{I_2} \cdots \sum_{i_N=1}^{I_N} x_{i_1 i_2 \ldots i_N}^2}$$

The *Weighted norm* of $\mathcal{X}$ for two tensors $\mathcal{X}$ and $\mathcal{W}$ is defined as follows:

$$\|\mathcal{X}\|_W = \|\mathcal{W} * \mathcal{X}\|$$

The *Khatri-Rao product* $\odot$ is defined as follows:

$$\boldsymbol{X} \odot \boldsymbol{Y} = [\boldsymbol{x}_1 \otimes \boldsymbol{y}_1 \;\; \boldsymbol{x}_2 \otimes \boldsymbol{y}_2 \ldots \boldsymbol{x}_K \otimes \boldsymbol{y}_K]$$

where size of matrices $\boldsymbol{X}$ and $\boldsymbol{Y}$ is $I \times K$ and $J \times K$ respectively. The symbol $\otimes$ is the Kronecker product.

The *Kronecker product* $\otimes$ is defined as follows:

$$\boldsymbol{X} \otimes \boldsymbol{Y} = \begin{pmatrix} x_{11}\boldsymbol{Y} & \cdots & x_{1n}\boldsymbol{Y} \\ \vdots & \ddots & \vdots \\ x_{m1}\boldsymbol{Y} & \cdots & x_{mn}\boldsymbol{Y} \end{pmatrix}$$

where $\boldsymbol{X}$ is an $m \times n$ matrix and $\boldsymbol{Y}$ is a $p \times q$ matrix, and the Kronecker product $\boldsymbol{X} \otimes \boldsymbol{Y}$ is the $mp \times nq$ block matrix.

The *Outer product* $\circ$ between two vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ is given by:

$$\boldsymbol{x} \circ \boldsymbol{y} = \boldsymbol{x}\boldsymbol{y}^T$$

where $\boldsymbol{x}$ and $\boldsymbol{y}$ are column vectors and their outer product gives rank-1 matrix.

Tensor mode-n unfolding, which is also called tensor matricization, is analogous to vectorizing a matrix. Mode-n unfolding of $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ re-arranges the elements of $\mathcal{X}$ to form a matrix $\mathbf{X}_{(n)} \in \mathbb{R}^{I_n \times I_1 I_2 \dots I_{n-1} I_{n+1} I_N}$, where $I_n I_{n+1} I_{n+2} \dots I_N I_1 I_2 \dots I_{n-1}$ is in a cyclic order.

The notation $[\![\mathbf{A}^{(1)}, \mathbf{A}^{(2)}, \dots, \mathbf{A}^{(N)}]\!]$ defines a tensor of size $\mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ whose elements are given by:

$$([\![\mathbf{A}^{(1)}, \mathbf{A}^{(2)}, \dots, \mathbf{A}^{(N)}]\!]_{i_1, i_2, \dots, i_n} = \sum_{r=1}^{R} \prod_{n=1}^{N} a_{i_n r}^{(n)}$$

for $i_n \epsilon \{1, \dots, I_n\}$, $n \epsilon \{1, \dots N\}$.

3.3 Missing data and its types

Missing data has been categorized into two types: 1) unstructured missing data 2) structured missing data. If the observed data in the original structure is missing randomly; it is categorized as unstructured missing data as shown in Fig. 3.2 (a). For example, samples of EMG data missing at random entries. However, if the data is missing in some consistent and structured way, it is termed as structured missing data. For example, some percentage of consecutive values of an EMG channel are missing either in the beginning, middle or end of data acquisition process/session as shown in Fig. 3.2 (b). This block of missing values is repeated randomly in other channels of EMG data.

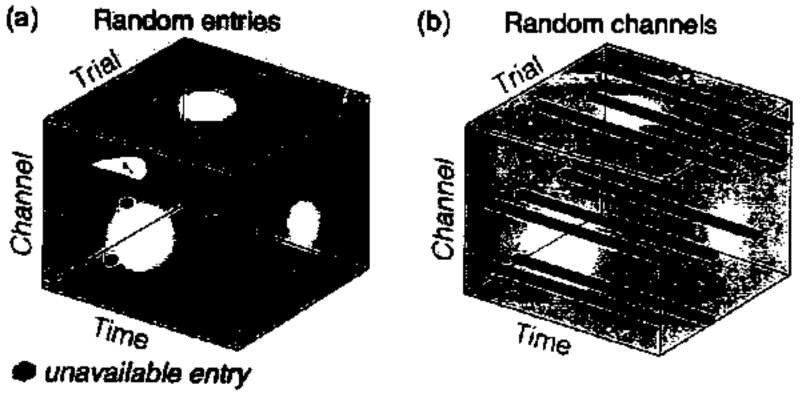

Fig. 3.2 Unstructured and structured missing data. a) Unstructured missing data with data missing at random instants b) Structured missing data with missing channels [57]

3.4 Tensor factorization

A tensor is a generalized matrix. A vector with order one can be termed as one-dimensional (1-D) tensor, whereas a matrix with order two can be termed as two-dimensional (2-D) tensor. N-dimensional tensor is represented as $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$, where the order of $\mathcal{X}$ is N (N > 2). A 3-D tensor is a higher order array with order N = 3. Mode is a term used to represent each dimension of a tensor. Hence for $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$, N is the mode where

I_N is the dimension in the Nth mode. For 3-D tensor where N = 3, $\mathcal{X}$ would have three dimensions I_1, I_2 and I_3, and $\mathcal{X}$ would be represented as $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$. Elementwise $\mathcal{X}$ would be written as $x_{i_1 i_2 i_3}$ where i_1, i_2 and i_3 are the entries in i_1^{th} row, i_2^{th} column and i_3^{th} tube as shown in Fig. 3.3. If an index is fixed in one of the modes and the indices vary in other two modes such division of data is called a slice. For example if i_1^{th} row of $\mathcal{X}$ is fixed and indices vary across i_2 and i_3 then it is a horizontal slice as shown in Fig. 3.4 (a). Likewise if i_2^{th} column of $\mathcal{X}$ is fixed and indices vary across i_1 and i_3 then it is a vertical slice as shown in Fig. 3.4 (b).

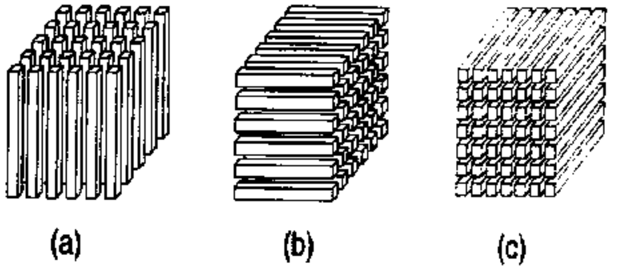

Fig. 3.3 (a) Columns (b) Rows (c) tubes [58]

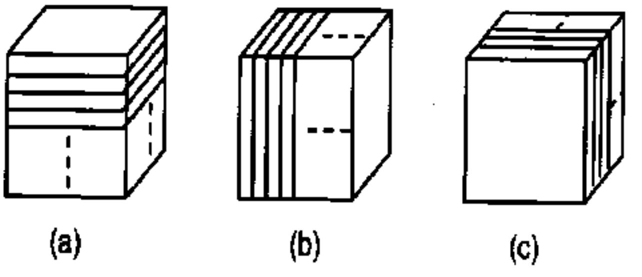

Fig. 3.4 a) Horizontal slices b) Vertical slices c) Frontal slices [58]

3.5 Signal Processing

3.5.1 NMF

The objective function for recovering missing values of the EMG data in the form of Matrix is given as:

$$f(\overline{X}) = \min_{\overline{X}} \|X - \overline{X}\|_F^2 \quad (3.1)$$

where $X \in \mathbb{R}^{m\times n}$ is the input matrix which contains EMG data with missing values and $\overline{X}$ is the reconstructed matrix obtained by minimizing the objective function in equation (3.1). To solve equation (3.1) using NMF [18], the objective function in equation (3.1) becomes:

$$f(P,Q) = \min_{P,Q} \|X - PQ\|_F^2 \quad (3.2)$$

where P and Q are $\mathbb{R}^{m\times k}$ and $\mathbb{R}^{k\times n}$ matrices, respectively."

In order to apply NMF to multidimensional input data, it is represented as a matrix **X** with dimensions time × channels. NMF decomposes the data of matrix **X** into two matrices P and Q, as mentioned above. Main objective of NMF is to find factor matrices P and Q that minimize the objective function in equation (3.2).

3.5.2 PARAFAC

The objective function for recovering missing values of the EMG data in the form of tensors is given as:

$$f(\overline{\mathcal{X}}) = \min_{\overline{\mathcal{X}}} \| \mathcal{X} - \overline{\mathcal{X}} \|_F^2 \quad (3.3)$$

where $\mathcal{X} \in \mathbb{R}^{I_1\times I_2\times \ldots \times I_N}$ is an order-Ninput tensor and assumed rank is R. $\mathcal{X}$ contains EMG data with missing values and $\overline{\mathcal{X}}$ is the reconstructed tensor obtained by minimizing the objective function. To solve equation (3.3), a standard tensor factorization is CP, which can be used to find the reconstructed tensor, then the objective function in equation (3.3) becomes:

$$f\left(\mathbf{A}^{(1)}\mathbf{A}^{(2)},\dots,\mathbf{A}^{(N)}\right)=\min_{\mathbf{A}^{(1)}\mathbf{A}^{(2)},\dots,\mathbf{A}^{(N)}}\frac{1}{2}\left\|\left(\mathcal{X}-[\![\mathbf{A}^{(1)},\mathbf{A}^{(2)},\dots,\mathbf{A}^{(N)}]\!]\right)\right\|_F^2 \tag{3.4}$$

where $\mathbf{A}^{(n)}$is factor matrix corresponding to n-th dimension, $[\![\mathbf{A}^{(1)},\mathbf{A}^{(2)},\dots,\mathbf{A}^{(N)}]\!]$ makes an order-N tensor equivalent to:

$$\overline{\mathcal{X}}\approx[\![\mathbf{A}^{(1)},\mathbf{A}^{(2)},\dots,\mathbf{A}^{(N)}]\!]\equiv\sum_{r=1}^{R}\mathbf{a}_r^{(1)}\circ\mathbf{a}_r^{(2)}\circ\cdots\circ\mathbf{a}_r^{(N)} \tag{3.5}$$

where $\mathbf{a}_r^{(n)}$is r-th column vector of $\mathbf{A}^{(n)}$factor matrix , and $n=1,2,\dots,N$. The sum of the outer products of vectors $\mathbf{a}_r^{(n)}$in equation (3.6) shows the CP decomposition as a sum of R *rank*-1 tensors to estimate a tensor. CP decomposition [2] is used to find the factor matrices of the input tensor. The tensor in equation (3.5) is an approximation proposed by CP/PARAFAC method [2] which is one of the standard methods for tensor factorization. In equation (3.5), a particular constraint is the value of R which is determined heuristically. CP tensor factorization method is modified to a weighted CP model which caters for the missing data recovery. Element wise, equation (3.5) can be written as:

$$\bar{x}_{i_1 i_2 \dots i_N}=\sum_r^R a_{i_1 r}a_{i_2 r}\dots a_{i_N r} \tag{3.6}$$

for $i_1=1,2,\dots,I_1, i_2=1,2,\dots,I_2,\cdots i_N=1,2,\dots,I_N$

In mode-n unfolded (matrix) form, equation (3.5) is represented as:

$$\overline{\mathbf{X}}_{(n)}=\mathbf{A}^{(n)}\left(\mathbf{A}^{(-n)}\right)^T \tag{3.7}$$

where

$$\mathbf{A}^{(-n)} = \mathbf{A}^{(N)} \odot \ldots \odot \mathbf{A}^{(n+1)} \odot \mathbf{A}^{(n-1)} \odot \ldots \odot \mathbf{A}^{(1)}$$

In unfolded form, the objective function to find mode-n factor matrices becomes:

$$f\left(\mathbf{A}^{(1)}\mathbf{A}^{(2)}, \ldots, \mathbf{A}^{(N)}\right) = \min_{\mathbf{A}^{(n)}} \frac{1}{2} \left\| \left(\mathbf{X}_{(n)} - \mathbf{A}^{(n)} \left(\mathbf{A}^{(N)} \odot \ldots \odot \mathbf{A}^{(n+1)} \odot \mathbf{A}^{(n-1)} \odot \ldots \odot \mathbf{A}^{(1)} \right)^T \right) \right\|_F^2$$

Literature provides many methods to compute CP decomposition to find a good approximation of original data such as alternating least squares (ALS) [59],[60], gradient descent (GD) [60] and enhanced line search [60] etc.

Our experiments show that conventional method such as CP decomposition only give comparable results to that of matrix factorization methods that even worsens when large amount of data is missing. To overcome this problem, we model CP factor matrices only from non-zero values of the input data. For this purpose, we multiply the input data with a weighting tensor $\mathcal{W}$ with size equal to the size of input data tensor $\mathcal{X}$ such that

$$w_{i_1 i_2 \ldots i_N} = \begin{cases} 1 \text{ if } x_{i_1 i_2 \ldots i_N} \text{is known} \\ 0 \text{ if } x_{i_1 i_2 \ldots i_N} \text{is missing} \end{cases}$$

for all $i_1 = 1,2, \ldots, I_1, i_2 = 1,2, \ldots, I_2, \cdots i_N = 1,2, \ldots, I_N$.

The weighted CP factorization of the EMG tensor yield factor matrices, which reconstruct the tensor using equation (3.7) to estimate the missing values.

3.5.3 CP-WOPT

CP-WOPT solves the problem of fitting the CP model to missing data by solving the following weighted least-squares objective function:

$$f_W\left(\mathbf{A}^{(1)}\mathbf{A}^{(2)},\ldots,\mathbf{A}^{(N)}\right)=\frac{1}{2}\left\|\left\{\left(\mathcal{X}-\left[\!\left[\mathbf{A}^{(1)}\mathbf{A}^{(2)},\ldots,\mathbf{A}^{(N)}\right]\!\right]\right)\mathcal{W}\right\}\right\|_{\mathcal{X}}^{2} \quad (3.8)$$

where $\mathcal{W}$ is tensor of the same size as $\mathcal{X}$, and its samples are defined as:

$$w_{ijk}=\begin{cases} 1 \text{ if } x_{ijk} \text{is known} \\ 0 \text{ if } x_{ijk} \text{is unknown} \end{cases} \quad (3.9)$$

for all $i = 1,\ldots,I$,$J = 1,\ldots,J$ and $k = 1,\ldots,K$.

For the sake of simplicity equation (3.8) is redefined as:

$$f_W\left(\mathbf{A}^{(1)}\mathbf{A}^{(2)},\ldots,\mathbf{A}^{(N)}\right)=\frac{1}{2}\left\|\mathcal{Y}-\mathcal{Z}\right\|^{2} \quad (3.10)$$

where

$$\mathcal{Y}=\mathcal{W}*\mathcal{X} \text{ and } \mathcal{Z}=\mathcal{W}*\left[\!\left[\mathbf{A}^{(1)}\mathbf{A}^{(2)},\ldots,\mathbf{A}^{(N)}\right]\!\right] \quad (3.11)$$

The gradient equation for the weighted case would be:

$$\frac{\partial f_W}{\partial \mathbf{A}^{(n)}}=\left(\mathbf{Z}^{(n)}-\mathbf{Y}^{(n)}\right)\mathbf{A}^{(-n)}, \quad (3.12)$$

for $n = 1,\ldots,N$.

The main objective is to find factor matrices $\mathbf{A}^{(n)} \in \mathbb{R}^{I_n \times R}$ for $n = 1,\ldots,N$ that minimize the weighted objective function in equation (3.10). Once gradients in equation (3.12) are known, any gradient-based optimization method can be used to solve the optimization problem. We use CP-WOPT [59] and the nonlinear conjugate gradient (NCG) as the optimization method with Hestenes-Stiefel updates [61]. The stopping conditions of both tensor based algorithms are based on the relative change in the function value f_W in equation (3.8) (set to 10^{-8}). The maximum number of iteration is set to 10^3 and the maximum number of function evaluations is set to 10^4. These choices are based on the values used in [2]. The brief methodology of CP-WOPT is summarized below:

Algorithm Methodology of CP-WOPT

Task: To find gradient matrices $G^{(n)}$ that minimize the weighted objective function in (6).

Input: $\mathcal{X}$ (Input tensor with missing values)

Output: $G^{(n)}$

Steps to compute $G^{(n)}$:

1. Compute $\mathcal{Y} = \mathcal{W} * \mathcal{X}$
2. Compute $\mathcal{Z} = \mathcal{W} * [\![A^{(1)}A^{(2)} \ldots, A^{(N)}]\!]$
3. Compute value of functions: $f = \frac{1}{2} - (\mathcal{Y}, \mathcal{Z}) + \frac{1}{2}\|\mathcal{Z}\|^2$
4. Compute $\mathcal{T} = \mathcal{Y} - \mathcal{Z}$

Repeat for $n = 1, \ldots, N$:

5. $G^{(n)} = -T_{(n)}A^{-(n)}$

3.6 Theoretical Framework

Missing data in EMG has been categorized into two types: 1) unstructured missing data 2) structured missing data. If the observed data in the original structure is missing randomly, then such a pattern of missing data is categorized as unstructured missing data. For example, samples of EMG data missing at random entries. However, if the data is missing in some consistent and structured way, it is termed as structured missing data. For example, 25% consecutive values of an EMG channel are missing either at the start, middle or end of data acquisition process/session. This block of missing values is repeated randomly in other channels of EMG data.

3.7 Experimental setup and procedure

3.7.1 Study one

Surface EMG signals of ten healthy subjects were acquired using six surface EMG electrodes. Three electrodes were placed on flexor and three electrodes on extensor muscles. The sampling frequency of surface EMG signals was 8 kHz, whereas we filtered it using bandpass filter of third order with bandwidths 20-500 Hz. Total of four-hand motions were performed by each subject: (1) hand open (2) hand close (3) pronation and (4) hand extension. Each hand motion was repeated four times per session, with a contraction and relaxation time of five seconds. Hence a single session span over a time interval of 400 seconds.

3.7.2 Study two

The experimental setup for this study mainly focuses on recovering missing data in iEMG signals. iEMG signals were collected by inserting six pairs of wires into the flexor carpi radialis, palmaris longus muscle, flexor digitorum superficialis, extensor carpi radialis longus, extensor digitorum, and extensor carpi ulnaris [3]. Fig. 3.6 shows placement of electrodes. Intramuscular EMG signals were filtered digitally with a third-order Butterworth bandpass filter of 100–900 Hz and sampled at 8 kHz. The 100-900 Hz range of bandpass filter was selected because useful frequency contents in iEMG lie within this range [3].

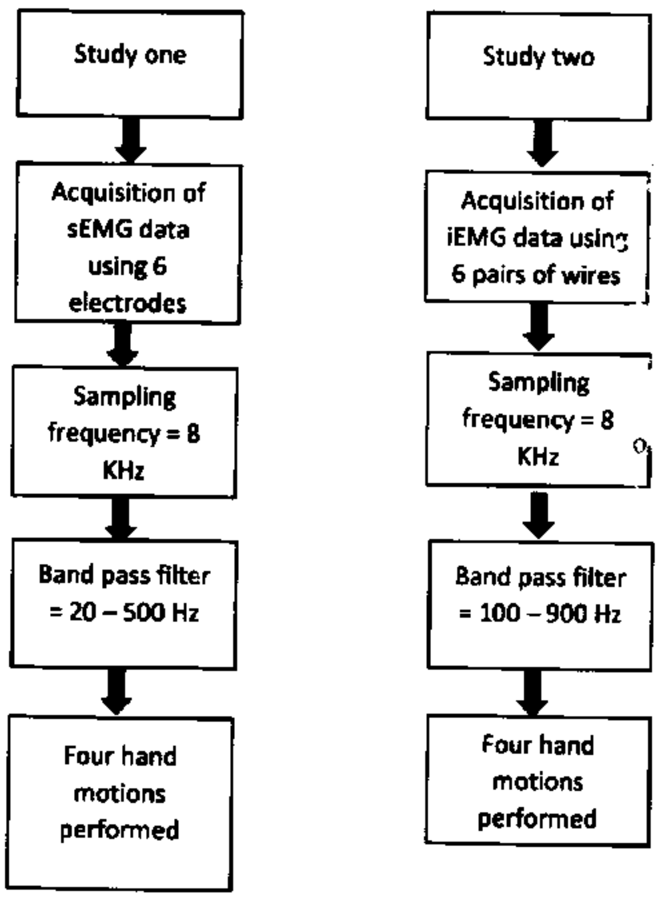

Fig. 3.5 Flowchart of two studies

Each of the ten subjects performed four hand motions in each experimental session: hand open, hand close, pronation and extend hand. For each subject, seven experimental sessions were conducted where each session was separated by 24 hours. Each hand movement was repeated four times with a contraction and relaxation time of 5 seconds per session. The order of the movements was selected randomly.

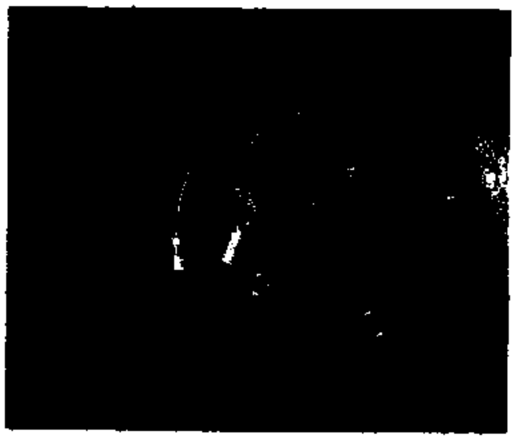

Fig. 3.6 Electrodes placement

3.8 Data Acquisition

3.8.1 Synthetic data

For the first study, a tensor of size $\mathbf{R}^{I\times J\times K}$ was generated comprising a number of true factors $R = 5$. To test the performance of different methods for the recovery of unstructured missing, a set of different sized data (such as $60 \times 50 \times 40, 120 \times 100 \times 80, 180 \times 150 \times 120$) was synthesized. For the case of structured missing data, the methods were tested on a dataset of size $120 \times 100 \times 80$. Factor matrices **A**, **B** and **C** were generated with sizes: $\mathbf{R}^{I\times R}, \mathbf{R}^{J\times R}$ and $\mathbf{R}^{K\times R}$ respectively. All the factor matrices were randomly chosen from $\mathcal{N}(0, 1)$ and then normalized every column to unit length. We then create the data tensor as:

$$\mathcal{X} = [\![\mathbf{A}, \mathbf{B}, \mathbf{C}]\!] + \eta \frac{[\mathcal{X}]}{[\mathcal{N}]} \quad (3.13)$$

Here $\mathcal{N}$ is a noise tensor (of the same size as $\mathcal{X}$) in which all samples were drawn from Gaussian i.i.d. distribution with mean zero and variance one. The term $[\![\mathbf{A}, \mathbf{B}, \mathbf{C}]\!]$ is a tensor being constructed from factor matrices **A**, **B** and **C** where η is noise parameter which has value 0.1.

3.8.2 Real-life EMG data

For sEMG data acquisition, six electrodes were used to collect sEMG signals on a single day. The movement-wise size of data was $320000 \times 6 \times 4$, which was down-sampled to $80000 \times 6 \times 4$. 80000 is the number of samples, 6 represents total number of electrodes/channels and 4 is total number of movements for which EMG data was collected. Data was down-sampled to reduce its size and computation time. Sampling factor was kept smaller to keep the data unaffected. After down-sampling EMG data, we normalized it between 0 and 1. Surface EMG data in the form of a tensor $\mathcal{X}$ can be viewed as $\mathcal{X} \in \mathbf{R}^{80000 \times 6 \times 4}$ for each of four movements. If we relate it with Fig. 3.1, then I = 80000, J = 8 and K = 4 where I, J and K represent samples of EMG data, total number of channels and total number of movements respectively.

Ten healthy subjects (all males; 25 ± 0.22 years (mean age ± SD)) and two transradial amputee subjects (all males, 34.8 ± 0.33 (mean age ± SD)) took part in the second study. These procedures were performed in accordance with the Declaration of Helsinki and approved by the local ethical committee of Riphah International University (approval no.: ref# Riphah/RCRS/REC/000121/20012016).

3.9 Statistical tests

A three-way ANOVA was used to assess which method had the least amount of RME. Three factors: methods (NMF, PARAFAC and CP-WOPT), Movement type (hand open, hand close, pronation and extend hand) and missing data percentage (10%, 20%, 30%, 40% and 50%) were used, post hoc pairwise comparisons were made using Tukey's HSD tests if required. Statistical significance was set at $P < 0.05$ for all comparisons.

3.10 Evaluation metric

Let $\mathbf{X}$ be the original data and $\bar{\mathbf{X}}$ be the estimated data produced by the matrix or tensor factorization methods. Then the Relative Mean Error (RME) is:

$$\text{RME} = \frac{\|(\mathbf{X} - \bar{\mathbf{X}})\|_F}{\|\mathbf{X}\|_F} \tag{3.14}$$

The best possible score is zero, i.e., the recovered data matches with original data completely.

3.11 Critique of methodology

Matrix and tensor factorization techniques are assessed to evaluate their performance to recover missing data for synthetic and real EMG data. For matrix and tensor factorization we applied NMF, and PARAFAC and CP-WOPT respectively. One of the reason for tensor factorization to outperform NMF is the arrangement of EMG data in a multidimensional way. This multidimensional arrangement of the data to constitute a tensor captures the global structure of observed data and models it efficiently by covering entire spatial and temporal dimension with an additional feature of multi-mode correlations. The performance of NMF and PARAFAC to recover missing data was almost the same as both the methods model, both known and unknown values.

Although PARAFAC is a tensor-based technique with the benefit of preserving the multi-way nature of data, yet its performance is comparable with NMF. The results reveal that CP-WOPT outperformed both NMF and PARAFAC to recover both unstructured and structured missing data. Usually, factorization methods find latent factors and then exploit those latent factors to predict the missing values. However, Matrix factorization based latent factors only capture two-dimensional linear relationships for estimating missing values, which can be improved if multi-linear relations are used. The main advantage of working through latent factors is that they let us take into account the information of the tensor explicitly by exploiting the multilinear interactions between obtained latent factors.

The main advantage of employing tensor factorization is that solution provided by it is unique [62]. Moreover, tensor factorization offers better computational capabilities and storage [63].

We divided the missing data into two categories: 1) unstructured missing data and 2) structured missing data. CP-WOPT gave promising results in recovering unstructured and structured (which is a more realistic assumption in Muscle-Computer Interface) missing data.

3.12 Summary

This chapter has described the methodology to employ matrix and tensor factorization methods on EMG data to recover missing data so that performance of prosthetic hand can be improved. Relevant notations and preliminaries are also mentioned which are used frequently in factorizations methods. Moreover, experimental setup to acquire data is discussed. Evaluation metric is mentioned to test the performance of factorization methods. Results of the performances of different factorization methods is discussed in Chapter 4.

Chapter 4.

Results and Discussion

4.1 Introduction

To the best of our knowledge, the problem of missing data in sEMG and iEMG signals has not been addressed so far. Therefore, for the first time, we will recover unstructured and structured missing data (random missing channels) in sEMG and iEMG signals and carry out detailed analyses and comparisons with matrix and tensor factorization methods. This is an extension of our work [33] where we recovered unstructured and structured missing data in surface EMG signals for ten subjects and data of only one day. We recovered missing data in iEMG signals as well which are collected for ten healthy and two amputee subjects over seven days. We employ NMF which is a matrix-based factorization method, and Canonical Polyadic Decomposition (CPD) & Canonical Polyadic-Weighted Optimization (CP-WOPT) which are both tensor-based factorization methods to recover structured missing data. They basically find the latent factors via high order factorization to estimate the missing data. We further extend our analysis for multiday iEMG data to test the performance of matrix and tensor factorization methods when the percentage of missing data is large.

4.2 Synthetic data performance

In Fig. 4.1, we compare the estimation performance of matrix and tensor-based factorization methods to recover unstructured missing data in the synthetic dataset for different proportions, e.g. 60%, 70%, 80%, 90% and 95%. In Fig. 4.2, we show the capability of different methods to recover structured missing data. Structured missing data is modelled by replacing entire 10, 20, 30, 40 and 50 columns (which are channels in case of EMG data) with zeroes. Performance of NMF, PARAFAC and CP-WOPT to recover unstructured and structured missing data is presented in Table 4.1 and Table 4.2 respectively.

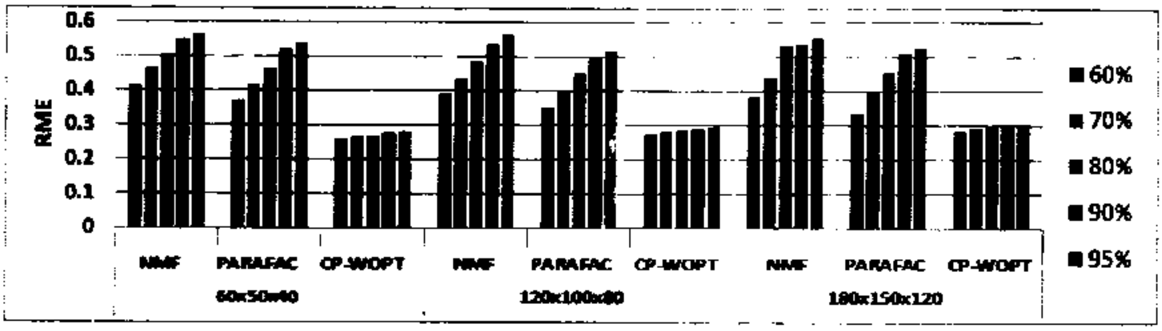

Fig. 4.1 RME of NMF, PARAFAC and CP-WOPT methods for 60%, 70%, 80%, 90% and 95% missing data for synthetic data of sizes 60×40×40, 120×100×80 and 180×150×120.

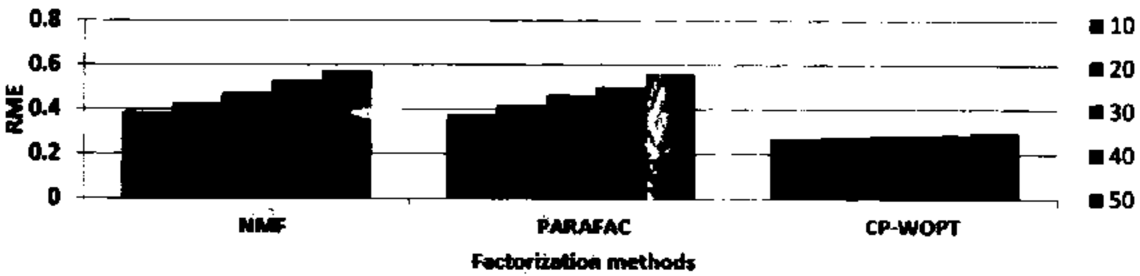

Fig. 4.2 RME of NMF, PARAFAC and CP-WOPT methods for 10, 20, 30, 40 and 50 columns missing in structured manner from synthetic data of size 120×100×80

4.3 Surface EMG data performance

Fig. 4.3 shows a segment of original EMG data with no missing values, the same EMG segment with unstructured missing values, and lastly the recovered EMG signal. A segment of the original EMG signal is shown in Fig. 4.3(a) with no missing values and it contains information of movement of muscle from a single channel. Fig. 4.3(b) shows the same EMG signal with unstructured missing values, which are the input signal to factorization methods. It can be seen in Fig. 4.3(b) that a lot of values with different amplitudes are replaced by zeroes to model unstructured missing data. Fig. 4.3(c) shows a recovered EMG signal when CP-WOPT is applied on the EMG signal of Fig. 4.3(b). It can be seen in Fig. 4.3(c) that all the missing values that were replaced by zeroes were successfully recovered with amplitudes around 0.48 ± 0.02. Fig. 4.4 illustrates a segment of EMG data with no missing values having same four movements where each movement exists at higher amplitudes. First half (with two movements) and second half (with two movements) is removed and then recovered from EMG data. In Fig. 4.4(a), EMG signal with no missing values is shown that has been obtained from a particular channel.

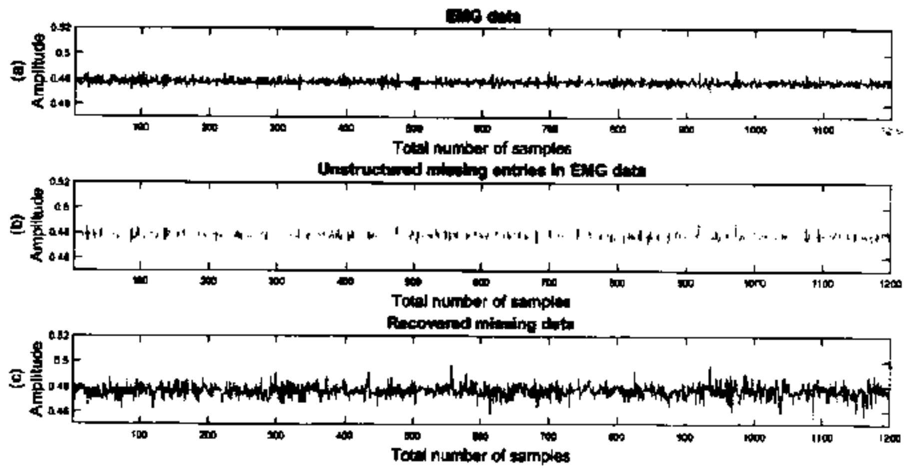

Fig. 4.3 (a) Original EMG data (b) Unstructured missing data (c) Recovered missing data by CP-WOPT.

The four epochs of higher amplitudes indicate execution of movement, however it can be seen that there is a very small difference between amplitudes of movement and no-movement (at rest) epochs. In Fig. 4.4 (b & d), first and second half (the worst case of removing 50% of data) of channel values is removed to model the structured missing data.

This simulates the scenario where data of two movements is completely missed. In Fig. 4.4(c & e), recovered signal by CP-WOPT is shown in which it can be seen that the difference between amplitudes of movement and no-movement epochs have increased which clearly differentiate epochs.

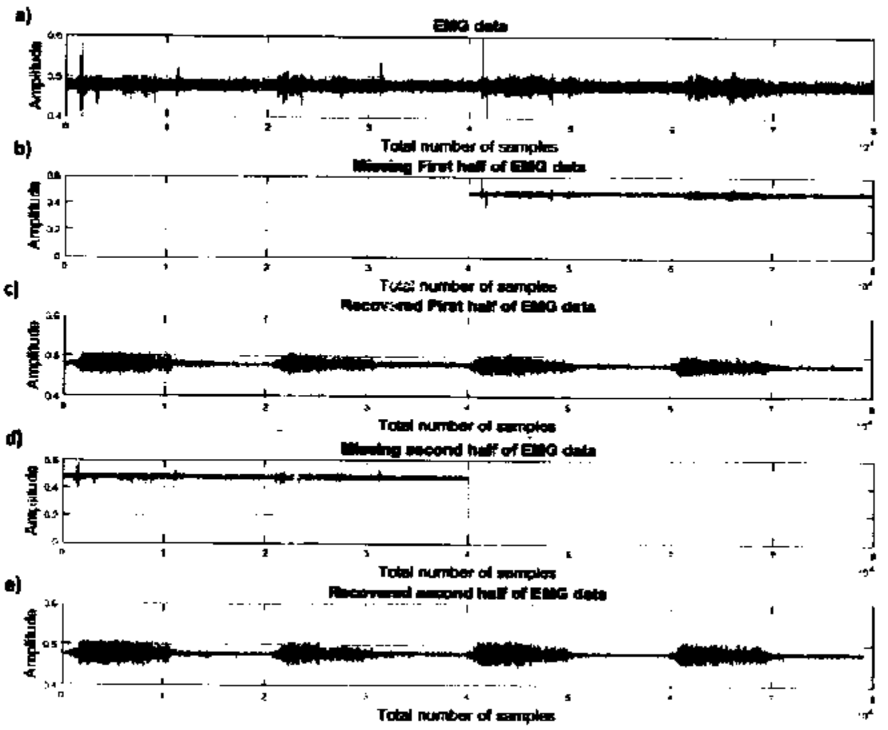

Fig. 4.4 (a) Original EMG data (b & d) First and Second half of a channel missing (c) & (e) Recovered missing channel by CP-WOPT.

Fig. 4.5 shows a comparison of three methods to recover unstructured missing data. There was a significant decrease ($P<0.05$) in the RME value with CP-WOPT as compared to PARAFAC and NMF across all four movements and different percentage of missing data. From each of four movements, we remove 60%, 70%, 80%, 90% and 95% data randomly in an unstructured manner.

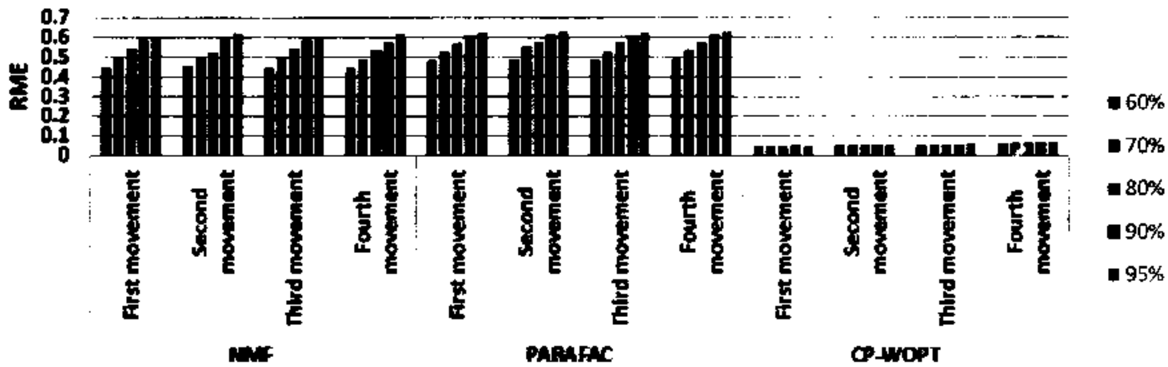

Fig. 4.5 RME for recovering 60%, 70%, 80%, 90% and 95% unstructured EMG missing data by NMF, PARAFAC and CP-WOPT.

Performance of NMF, PARAFAC and CP-WOPT to recover unstructured missing data in sEMG signals is presented in Table 4.3, Table 4.4 and Table 4.5 respectively. Whereas, performance of NMF, PARAFAC and CP-WOPT to recover structured missing data in sEMG signals is presented in Table 4.6, Table 4.7 and Table 4.8.

Table 4.1 Performance of NMF, PARAFAC and CP-WOPT on synthetic data to recover unstructured missing data of various sizes

	60x50x40			120x100x80			180x150x120		
Missing data (%)	**NMF**	**PARAFAC**	**CP-WOPT**	**NMF**	**PARAFAC**	**CP-WOPT**	**NMF**	**PARAFAC**	**CP-WOPT**
60%	0.4152	0.371	0.2612	0.3921	0.3521	0.2741	0.3856	0.3352	0.284
70%	0.4659	0.4171	0.2681	0.4354	0.3954	0.2812	0.4397	0.3975	0.2928
80%	0.506	0.4649	0.2701	0.4859	0.4526	0.2854	0.5348	0.4557	0.2991
90%	0.5494	0.5209	0.2751	0.5348	0.4987	0.2912	0.5374	0.5087	0.301
95%	0.5676	0.5414	0.2811	0.5641	0.5174	0.2954	0.5562	0.5286	0.3059

Table 4.2: Performance of NMF, PARAFAC and CP-WOPT on synthetic data to recover structured missing data of various sizes

	Number of missing channels				
	10	**20**	**30**	**40**	**50**
NMF	0.3872	0.4306	0.4731	0.5308	0.5742
PARAFAC	0.3787	0.4251	0.4639	0.5031	0.5619
CP-WOPT	0.271	0.2777	0.2812	0.2839	0.2987

Table 4.3: Performance of NMF on sEMG data to recover unstructured missing data of various percentages

NMF				
Mean	First movement	Second movement	Third movement	Fourth movement
60%	0.44704	0.4532	0.4463	0.4454
70%	0.4966	0.5053	0.4993	0.4893
80%	0.5456	0.5226	0.546	0.5367
90%	0.5866	0.593	0.5869	0.5719
95%	0.6047	0.6216	0.6048	0.6192

Table 4.4: Performance of PARAFAC on sEMG data to recover unstructured missing data of various percentages

PARAFAC				
Mean	**First movement**	**Second movement**	**Third movement**	**Fourth movement**
60%	0.483	0.4887	0.48442	0.4904
70%	0.53	0.5545	0.5304	0.536
80%	0.5711	0.5764	0.5726	0.578
90%	0.608	0.6145	0.61	0.6153
95%	0.6261	0.6312	0.6275	0.633

Table 4.5: Performance of CP-WOPT on sEMG data to recover unstructured missing data of various percentages

CP-WOPT				
Mean	**First movement**	**Second movement**	**Third movement**	**Fourth movement**
60%	0.0476	0.054	0.05435	0.06263
70%	0.0477	0.05361	0.05515	0.06343
80%	0.04786	0.0538	0.05445	0.06541
90%	0.049	0.056	0.05589	0.0674
95%	0.05	0.058	0.0587	0.071

Table 4.6: Performance of NMF on sEMG data to recover structured missing data of various percentages

NMF										
Mean	**1st half**					**2nd half**				
1st Movement	0.15925	0.19825	0.26165	0.29435	0.34085	0.17005	0.21305	0.25385	0.30935	0.35055
2nd Movement	0.14425	0.20745	0.27225	0.31565	0.35195	0.19985	0.20985	0.26985	0.29985	0.34985
3rd Movement	0.15265	0.22045	0.26765	0.31595	0.35315	0.16795	0.18995	0.24005	0.27485	0.29985
4th Movement	0.15265	0.22045	0.26765	0.31595	0.35315	0.179283	0.204283	0.254583	0.294683	0.333417

Table 4.7: Performance of PARAFAC on sEMG data to recover structured missing data of various percentages

PARAFAC										
Mean	**1st half**					**2nd half**				
1st Movement	0.15865	0.22405	0.27405	0.31635	0.35355	0.17475	0.22395	0.27425	0.31645	0.35375
2nd Movement	0.15965	0.22475	0.27485	0.31715	0.35435	0.17543	0.22425	0.27435	0.31675	0.35425
3rd Movement	0.15995	0.22495	0.27505	0.31725	0.35445	0.17485	0.22309	0.27405	0.3156	0.3523
4th Movement	0.15885	0.22385	0.27415	0.31645	0.35875	0.17105	0.22695	0.27915	0.33595	0.35825

Table 4.8: Performance of CP-WOPT on sEMG data to recover structured missing data of various percentages

CP-WOPT										
Mean	**1st half**					**2nd half**				
1st Movement	0.04405	0.04545	0.04985	0.06125	0.06415	0.04035	0.05125	0.05345	0.05855	0.09315
2nd Movement	0.03795	0.04805	0.04895	0.06725	0.08785	0.04115	0.06645	0.10525	0.10575	0.11995
3rd Movement	0.03175	0.03765	0.03785	0.03805	0.03815	0.0336	0.03585	0.04555	0.05235	0.07235
4th Movement	0.03855	0.04625	0.04755	0.04975	0.09085	0.04405	0.04655	0.05115	0.05365	0.07145

In Fig. 4.6, results are shown when NMF, PARAFAC and CP-WOPT are applied, respectively, to recover structured missing data. Results clearly show that CP-WOPT outperformed PARAFAC and NMF in recovering structured even for the extreme case when half of the channel data is missing. In structured missing data, we gradually increased the proportion of missing data from 10% to 50%. Removing 10% data from first half means data removal of first 10% samples from all six channels of particular movement whereas removing 50% data means data removal of first 50% samples (as shown in Fig. 4.4(b)) from all channels. Likewise, removing 10% data from second half means data removal of last 10% of samples from all six channels of particular movement, whereas removing 50% means data removal of last 50% of samples (as shown in Fig. 4.4(d)).

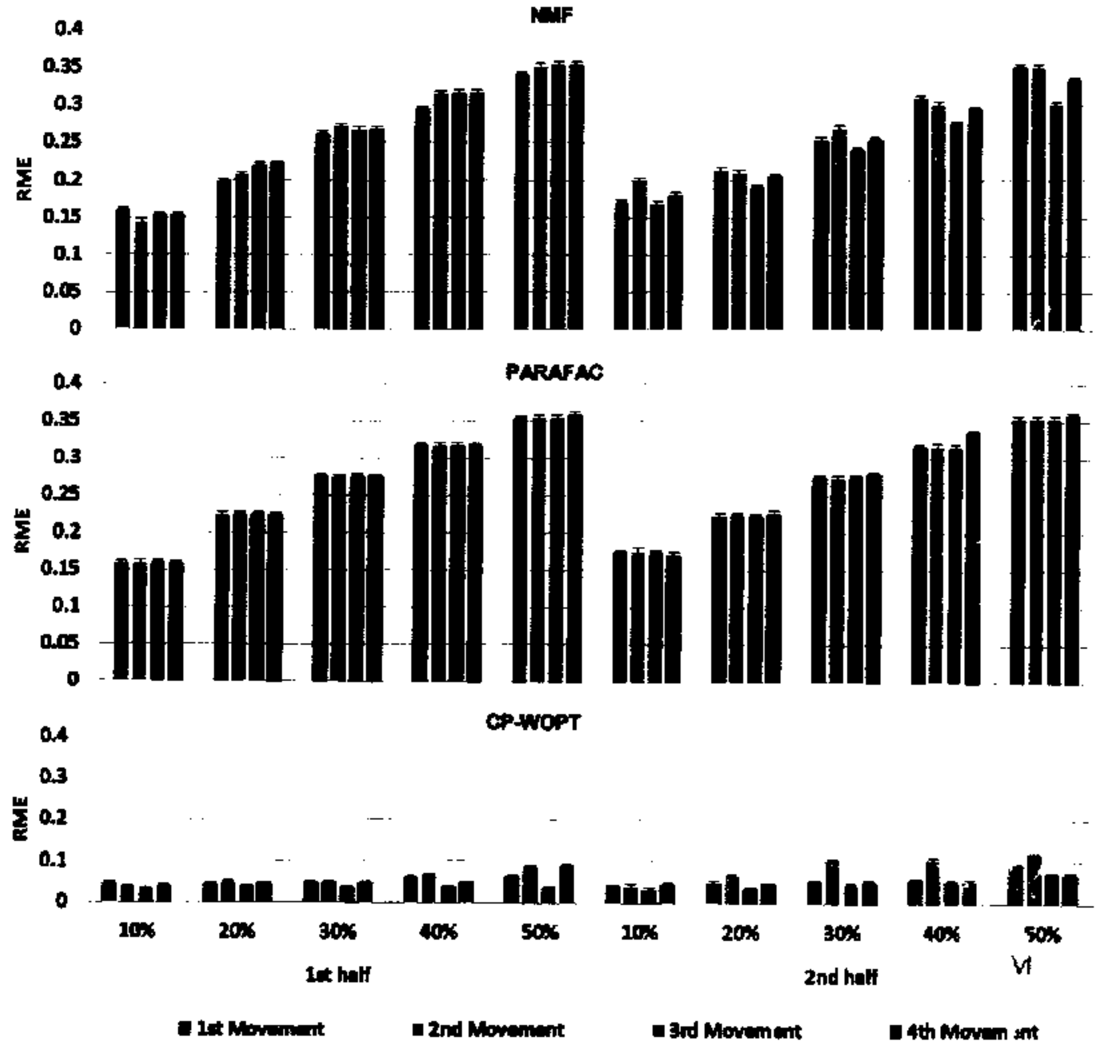

Fig. 4.6 RME for recovering structured missing samples from first and second half of real EMG data by NMF, PARAFAC and CP-WOPT.

In Fig. 4.7, we show computational complexity of NMF, PARAFAC and CP-WOPT. It can be seen that CP-WOPT takes slightly more time than NMF and PARAFAC to estimate 10%, 20%, 30%, 40% and 50% structured missing values to produce the reconstructed EMG data. In benchmark paper, EMG data of only single subject was acquired and 80% missing data was recovered. It is unclear whether the missing data was unstructured or structured. Moreover, the collected data was single day sEMG only and iEMG wasn't explored.

4.4 Missing iEMG performance

4.4.1 PERFORMANCE OF NMF

Fig. 4.8 (a) and Fig. 4.9 (a) show the performance of NMF to recover missing data in each of the four movements from day 1 to day 7 for healthy and amputee subjects respectively.

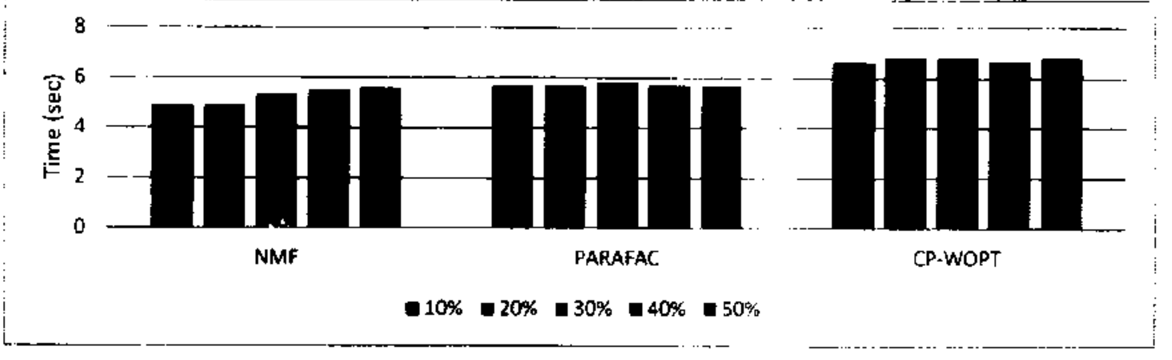

Fig. 4.7 Comparison of computational time of NMF, PARAFAC and CP-WOPT.

It can be seen in Fig.4.8 (a) and Fig. 4.9 (a) for day 1 that as the percentage of missing data from the first half of movement one increases from 10% to 50%, the RME also increases from 0.17 to 0.29 indicating degradation in performance of NMF to recover data. The performance of NMF is almost the same for the other three movements where RME increases (up to 0.29) with the increase in the percentage of missing data from 10% to 50%. The performance of NMF in terms of RME is almost similar for the same percentage of missing data for the second half of the movements as well. If we further analyze Fig. 4.8 (a) and Fig. 4.9 (a) for the case of missing data for day 1 to 4 and for each of the four movements, it can be seen that for the first half when the percentage of missing data is 10% the corresponding value of RME is 0.22 which increases to 0.35 when missing data increase to 50%. The performance of NMF is again almost same for second half. Moreover, if we analyze the case of missing data from 1 to 7 seven days and for each of the four movements, it can be seen that for first half when percentage of missing data is 10% the corresponding value of RME is much worse i.e. 0.22 which increases to 0.4 when missing

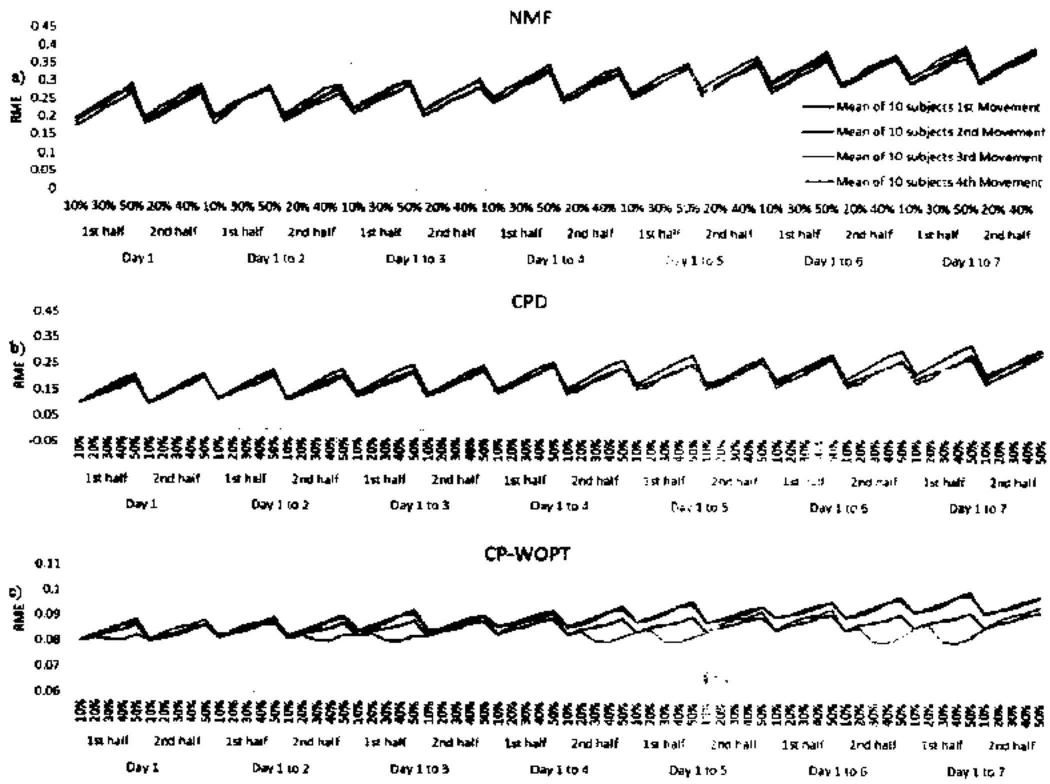

Fig 4.8 Movement-wise performance of (a) NMF, (b) CPD and (c) CP-WOPT, to recover missing data on healthy subjects for increasing percentage (10% to 50%) of missing iEMG data in both halves for increasing number of days (1 to 7).

data increase to 50%. The performance of NMF is again almost similar for the second half. If we analyze Fig. 4.8 (a) and Fig. 4.9 (a) as a whole, it can be seen that as the total amount of missing data increase from day 1 to 7, overall RME also increases from 0.17 to 0.4 with increase in percentage of missing data for either half. So, it can be seen that for small amount of missing data in small amount of observed data (day 1 only) NMF performs better as compared to increased amount of missing data and observed data for seven days where its performance degrades considerably. Performance of NMF to recover structured missing data of various percentages for day one to seven on iEMG data of healthy subjects has been presented from Table 4.9 to Table 4.15 respectively. Performance of NMF to recover structured missing data of various percentages for day one to seven on iEMG data of amputee subjects has been presented from Table 4.16 to Table 4.22 respectively.

Table 4.9: Performance of NMF on iEMG data of healthy subjects to recover structured missing data of various percentages for day one

Day 1											
Mean of 10 subject		1st half					2nd half				
		10%	20%	30%	40%	50%	10%	20%	30%	40%	50%
	1st Movement	0.197	0.216	0.239	0.265	0.279	0.192	0.213	0.237	0.253	0.28
	2nd Movement	0.196	0.221	0.247	0.269	0.273	0.2	0.229	0.247	0.272	0.291
	3rd Movement	0.176	0.196	0.224	0.243	0.266	0.184	0.201	0.225	0.245	0.267
	4th Movement	0.186	0.212	0.235	0.26	0.293	0.185	0.211	0.233	0.259	0.291

Table 4.10: Performance of NMF on iEMG data of healthy subjects to recover structured missing data of various percentages for day two

Day 2											
Mean of 10 subject		1st half					2nd half				
		10%	20%	30%	40%	50%	10%	20%	30%	40%	50%
	1st Movement	0.203	0.223	0.248	0.261	0.281	0.201	0.219	0.244	0.263	0.279
	2nd Movement	0.199	0.232	0.25	0.268	0.284	0.207	0.232	0.256	0.281	0.287

	3rd Movement	0.181	0.212	0.248	0.262	0.28	0.189	0.208	0.233	0.246	0.267
	4th Movement	0.199	0.217	0.247	0.263	0.287	0.195	0.22	0.238	0.26	0.292

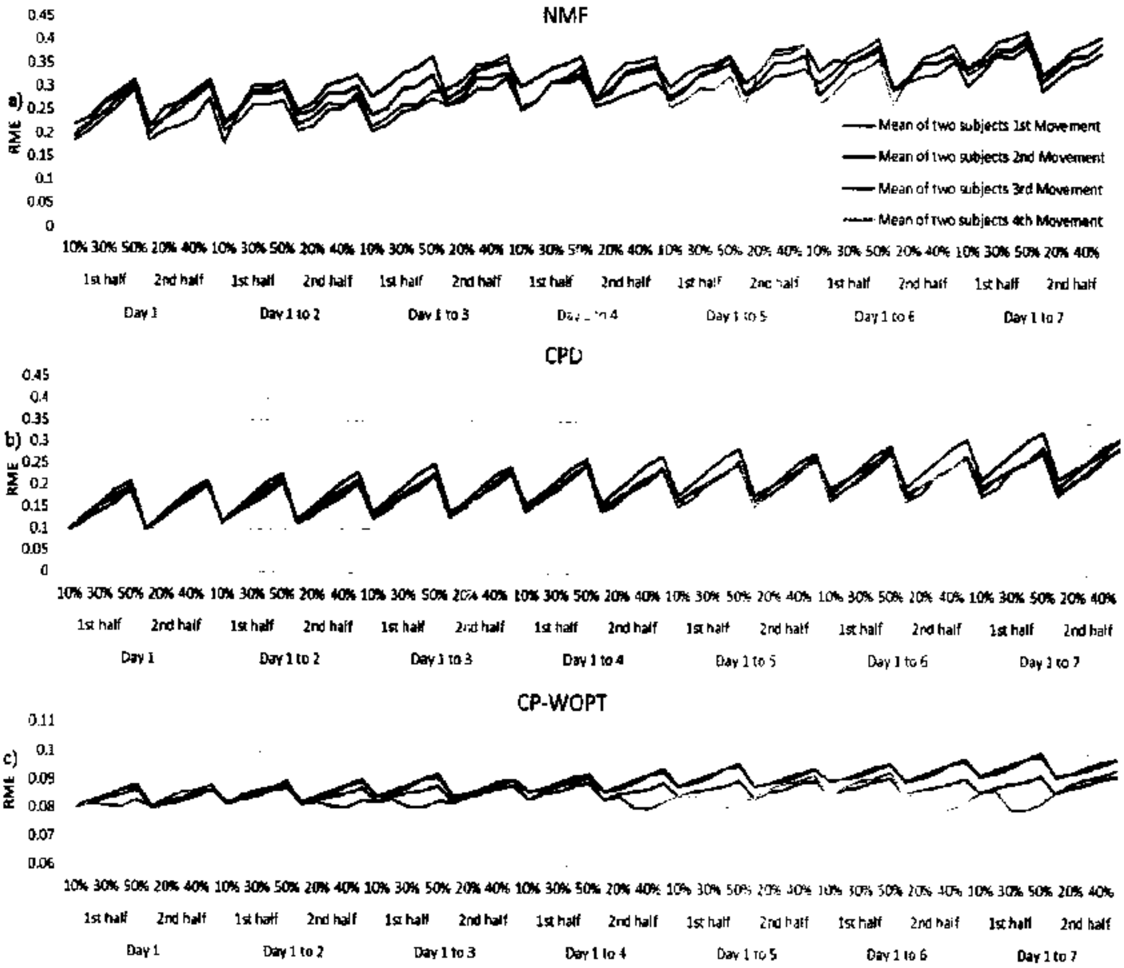

Fig. 4.9 Movement-wise performance of (a) NMF, (b) CPD and (c) CP WOPT, to recover missing data on amputee subjects for increasing percentage (10% to 50%) of missing iEMG data in both halves for increasing number of days (1 to 7).

Table 4.11: Performance of NMF on iEMG data of healthy subjects to recover structured missing data of various percentages for day three

Mean of 10 subject		Day 3									
		1st half					2nd half				
		10%	20%	30%	40%	50%	10%	20%	30%	40%	50%
	1st Movement	0.224	0.24	0.264	0.285	0.302	0.205	0.224	0.249	0.262	0.283
	2nd Movement	0.225	0.248	0.269	0.292	0.299	0.218	0.244	0.264	0.284	0.301
	3rd Movement	0.211	0.229	0.258	0.271	0.29	0.206	0.221	0.25	0.265	0.284
	4th Movement	0.211	0.231	0.25	0.266	0.292	0.2188	0.2386	0.2654	0.2822	0.308

Table 4.12: Performance of NMF on iEMG data of healthy subjects to recover structured missing data of various percentages for day tour

Mean of 10 subject		Day 4									
		1st half					2nd half				
		10%	20%	30%	40%	50%	10%	20%	30%	40%	50%
	1st Movement	0.242	0.263	0.289	0.306	0.332	0.246	0.264	0.289	0.308	0.324
	2nd Movement	0.254	0.282	0.302	0.324	0.346	0.255	0.278	0.302	0.327	0.333
	3rd Movement	0.248	0.264	0.295	0.313	0.334	0.242	0.261	0.286	0.299	0.32
	4th Movement	0.24	0.263	0.2898	0.30252	0.328	0.25	0.2786	0.2964	0.3142	0.342

Table 4.13: Performance of NMF on iEMG data of healthy subjects to recover structured missing data of various percentages for day five

Day 5											
Mean of 10 subject		1st half					2nd half				
		10%	20%	30%	40%	50%	10%	20%	30%	40%	50%
	1st Movement	0.254	0.272	0.297	0.32	0.34	0.27	0.292	0.318	0.331	0.353
	2nd Movement	0.268	0.292	0.314	0.334	0.343	0.281	0.309	0.33	0.352	0.37
	3rd Movement	0.268	0.287	0.316	0.331	0.352	0.266	0.283	0.313	0.328	0.348
	4th Movement	0.260054	0.281	0.29986	0.31734	0.34695	0.261	0.2926	0.3174	0.3352	0.363

Table 4.14: Performance of NMF on iEMG data of healthy subjects to recover structured missing data of various percentages for day six

Day 6											
Mean of 10 subject		1st half					2nd half				
		10%	20%	30%	40%	50%	10%	20%	30%	40%	50%
	1st Movement	0.272	0.293	0.319	0.336	0.362	0.287	0.305	0.33	0.349	0.365
	2nd Movement	0.299	0.327	0.338	0.355	0.377	0.289	0.313	0.337	0.362	0.368
	3rd Movement	0.284	0.3	0.331	0.349	0.37	0.296	0.316	0.344	0.356	0.376
	4th Movement	0.293	0.315	0.342	0.363	0.389	0.287	0.3156	0.3334	0.3512	0.379

Table 4.15: Performance of NMF on iEMG data of healthy subjects to recover structured missing data of various percentages for day seven

Day 7											
Mean of 10 subject		1st half					2nd half				
		10%	20%	30%	40%	50%	10%	20%	30%	40%	50%
	1st Movement	0.295	0.314	0.34	0.363	0.383	0.299	0.32	0.345	0.358	0.382
	2nd Movement	0.296	0.32	0.331	0.356	0.366	0.305	0.333	0.354	0.376	0.395
	3rd Movement	0.315	0.334	0.365	0.38	0.401	0.304	0.321	0.349	0.364	0.386
	4th Movement	0.30898	0.33	0.351	0.368	0.395	0.302	0.3242	0.3498	0.3674	0.396

Table 4.16: Performance of NMF on iEMG data of amputee subjects to recover structured missing data of various percentages for day one

Day 1											
Mean of 2 subject		1st half					2nd half				
		10%	20%	30%	40%	50%	10%	20%	30%	40%	50%
	1st Movement	0.195	0.22	0.245	0.275	0.305	0.215	0.23	0.265	0.275	0.305
	2nd Movement	0.22	0.235	0.27	0.29	0.315	0.215	0.255	0.265	0.29	0.315
	3rd Movement	0.185	0.205	0.235	0.26	0.295	0.185	0.205	0.215	0.23	0.275
	4th Movement	0.2	0.225	0.265	0.28	0.315	0.2	0.23	0.25	0.28	0.305

Table 4.17: Performance of NMF on iEMG data of amputee subjects to recover structured missing data of various percentages for day two

Day 2											
Mean of 2 subject		**1st half**					**2nd half**				
		10%	**20%**	**30%**	**40%**	**50%**	**10%**	**20%**	**30%**	**40%**	**50%**
	1st Movement	0.22	0.245	0.285	0.285	0.295	0.24	0.25	0.285	0.285	0.3
	2nd Movement	0.225	0.25	0.3	0.3	0.31	0.25	0.265	0.305	0.315	0.325
	3rd Movement	0.18	0.245	0.29	0.295	0.31	0.22	0.235	0.265	0.26	0.275
	4th Movement	0.205	0.225	0.26	0.26	0.27	0.205	0.215	0.25	0.255	0.285

Table 4.18: Performance of NMF on iEMG data of amputee subjects to recover structured missing data of various percentages for day three

Day 3											
Mean of 3 subject		**1st half**					**2nd half**				
		10%	**20%**	**30%**	**40%**	**50%**	**10%**	**20%**	**30%**	**40%**	**50%**
	1st Movement	0.24	0.255	0.295	0.3	0.325	0.26	0.28	0.32	0.32	0.33
	2nd Movement	0.28	0.3	0.33	0.34	0.365	0.275	0.295	0.335	0.345	0.355
	3rd Movement	0.205	0.22	0.245	0.255	0.275	0.26	0.27	0.295	0.295	0.32
	4th Movement	0.215	0.235	0.26	0.26	0.29	0.295	0.315	0.347	0.351	0.37

Table 4.19: Performance of NMF on iEMG data of amputee subjects to recover structured missing data of various percentages for day four

Mean of 4 subjects		Day 4									
		1st half					2nd half				
		10 %	20 %	30 %	40 %	50 %	10 %	20 %	30 %	40 %	50 %
	1st Movement	0.255	0.27	0.31	0.32	0.345	0.275	0.295	0.33	0.335	0.345
	2nd Movement	0.3	0.32	0.34	0.35	0.365	0.275	0.29	0.335	0.34	0.35
	3rd Movement	0.255	0.27	0.31	0.315	0.33	0.26	0.27	0.285	0.295	0.31
	4th Movement	0.25	0.27	0.31	0.31	0.32	0.275	0.323	0.352	0.356	0.365

Table 4.20: Performance of NMF on iEMG data of amputee subjects to recover structured missing data of various percentages for day five

Mean of 2 subjects		Day 5									
		1st half					2nd half				
		10%	20 %	30%	40%	50%	10 %	20 %	30 %	40 %	50 %
	1st Movement	0.275	0.295	0.33	0.34	0.365	0.285	0.315	0.355	0.355	0.37
	2nd Movement	0.3	0.33	0.345	0.35	0.37	0.31	0.335	0.38	0.385	0.395
	3rd Movement	0.28	0.3	0.325	0.335	0.35	0.285	0.3	0.325	0.33	0.34
	4th Movement	0.2577	0.275	0.2993	0.2967	0.32475	0.265	0.338	0.372	0.376	0.395

Table 4.21: Performance of NMF on iEMG data of amputee subjects to recover structured missing data of various percentages for day six

Day 6											
Mean of 2 subjects		**1st half**					**2nd half**				
		10 %	**20 %**	**30 %**	**40 %**	**50 %**	**10 %**	**20 %**	**30 %**	**40 %**	**50 %**
	1st Movement	0.285	0.315	0.355	0.365	0.38	0.295	0.32	0.355	0.355	0.37
	2nd Movement	0.335	0.36	0.355	0.37	0.39	0.295	0.325	0.365	0.375	0.395
	3rd Movement	0.31	0.335	0.37	0.385	0.405	0.295	0.315	0.325	0.33	0.355
	4th Movement	0.265	0.29	0.325	0.34	0.36	0.265	0.323	0.352	0.356	0.375

Table 4.22: Performance of NMF on iEMG data of amputee subjects to recover structured missing data of various percentages for day seven

Day 7											
Mean of 2 subjects		**1st half**					**2nd half**				
		10%	**20 %**	**30 %**	**40 %**	**50 %**	**10 %**	**20 %**	**30 %**	**40 %**	**50 %**
	1st Movement	0.305	0.335	0.375	0.385	0.41	0.295	0.32	0.345	0.355	0.375
	2nd Movement	0.34	0.36	0.37	0.365	0.39	0.315	0.345	0.38	0.395	0.41
	3rd Movement	0.345	0.365	0.4	0.41	0.42	0.33	0.35	0.365	0.37	0.395
	4th Movement	0.3299	0.35	0.385	0.38	0.4	0.315	0.338	0.367	0.371	0.395

4.4.2 PERFORMANCE OF CPD

Fig. 4.8(b) and Fig. 4.9(b) shows the performance of CPD in terms to recover missing data. It can be seen for both halves of Day 1 that as the percentage of missing data increases from 10% to 50%, RME also increase from 0.1 to 0.2 whereas for Day 1 to 7, RME increase from 0.19 to 0.33 for healthy subjects and 0.1 to 0.31 for amputee subjects indicating poor performance of CPD. Performance of CPD to recover structured missing data of various percentages for day one to seven on iEMG data of healthy subjects has been presented from Table 4.23 to Table 4.29 respectively. Performance of CPD to recover structured missing data of various percentages for day one to seven on iEMG data of amputee subjects has been presented from Table 4.30 to Table 4.36 respectively.

Table 4.23: Performance of CPD on iEMG data of healthy subjects to recover structured missing data of various percentages for day one

Day 1											
Mean of 10 subject		1st half					2nd half				
		10 %	20 %	30 %	40 %	50 %	10 %	20 %	30 %	40 %	50 %
	1st Movement	0.0995	0.1195	0.1395	0.1595	0.1895	0.1	0.13	0.16	0.19	0.21
	2nd Movement	0.0995	0.1295	0.1595	0.1895	0.2095	0.1002	0.1202	0.1502	0.1702	0.2002
	3rd Movement	0.1	0.118	0.158	0.178	0.198	0.0999	0.1299	0.1499	0.1799	0.2099
	4th Movement	0.1004	0.1304	0.1504	0.1704	0.1984	0.0993	0.1193	0.1493	0.1793	0.1993

Table 4.24: Performance of CPD on iEMG data of healthy subjects to recover structured missing data of various percentages for day two

Day 2											
Mean of 10 subject		1st half					2nd half				
		10 %	20 %	30 %	40 %	50 %	10 %	20 %	30 %	40 %	50 %
	1st Movement	0.1165	0.1465	0.1765	0.2065	0.2265	0.117	0.137	0.157	0.177	0.207
	2nd Movement	0.1195	0.1395	0.1595	0.1795	0.2095	0.1202	0.1502	0.1802	0.2102	0.2302
	3rd Movement	0.114	0.144	0.164	0.194	0.224	0.1139	0.1319	0.1719	0.1919	0.2119
	4th Movement	0.1154	0.1354	0.1654	0.1954	0.2154	0.1143	0.1443	0.1643	0.1843	0.2123

Table 4.25: Performance of CPD on iEMG data of healthy subjects to recover structured missing data of various percentages for day three

Day 3											
Mean of 10 subject		1st half					2nd half				
		10 %	20 %	30 %	40 %	50 %	10 %	20 %	30 %	40 %	50 %
	1st Movement	0.1335	0.1535	0.1735	0.1935	0.2235	0.134	0.164	0.194	0.224	0.244
	2nd Movement	0.1395	0.1695	0.1995	0.2295	0.2495	0.1402	0.1602	0.1802	0.2002	0.2302
	3rd Movement	0.128	0.146	0.186	0.206	0.226	0.1279	0.1579	0.1779	0.2079	0.2379
	4th Movement	0.1304	0.1604	0.1804	0.2004	0.2284	0.1293	0.1493	0.1793	0.2093	0.2293

Table 4.26: Performance of CPD on iEMG data of healthy subjects to recover structured missing data of various percentages for day four

Day 4											
Mean of 10 subject		1st half					2nd half				
		10 %	20 %	30 %	40 %	50 %	10 %	20 %	30 %	40 %	50 %
	1st Movement	0.1505	0.1805	0.2105	0.2405	0.2605	0.151	0.171	0.191	0.211	0.241
	2nd Movement	0.1595	0.1795	0.1995	0.2195	0.2495	0.1602	0.1902	0.2202	0.2502	0.2702
	3rd Movement	0.142	0.172	0.192	0.222	0.252	0.1419	0.1599	0.1999	0.2199	0.2399
	4th Movement	0.1454	0.1654	0.1954	0.2254	0.2454	0.1443	0.1743	0.1943	0.2143	0.2423

Table 4.27: Performance of CPD on iEMG data of healthy subjects to recover structured missing data of various percentages for day six

Day 5											
Mean of 10 subject		1st half					2nd half				
		10 %	20 %	30 %	40 %	50 %	10 %	20 %	30 %	40 %	50 %
	1st Movement	0.1675	0.1875	0.2075	0.2275	0.2575	0.168	0.198	0.228	0.258	0.278
	2nd Movement	0.1795	0.2095	0.2395	0.2695	0.2895	0.1802	0.2002	0.2202	0.2402	0.2702
	3rd Movement	0.156	0.174	0.214	0.234	0.254	0.1559	0.1859	0.2059	0.2359	0.2659
	4th Movement	0.1604	0.1904	0.2104	0.2304	0.2584	0.1593	0.1793	0.2093	0.2393	0.2593

Table 4.28: Performance of CPD on iEMG data of healthy subjects to recover structured missing data of various percentages for day six

Day 6											
Mean of 10 subject		1st half					2nd half				
		10%	20%	30%	40%	50%	10%	20%	30%	40%	50%
	1st Movement	0.1845	0.2145	0.2445	0.2745	0.2945	0.185	0.205	0.225	0.245	0.275
	2nd Movement	0.1995	0.2195	0.2395	0.2595	0.2895	0.2002	0.2302	0.2602	0.2902	0.3102
	3rd Movement	0.17	0.2	0.22	0.25	0.28	0.1699	0.1879	0.2279	0.2479	0.2679
	4th Movement	0.1754	0.1954	0.2254	0.2554	0.2754	0.1743	0.2043	0.2243	0.2443	0.2723

Table 4.29: Performance of CPD on iEMG data of healthy subjects to recover structured missing data of various percentages for day seven

Day 7											
Mean of 10 subject		1st half					2nd half				
		10%	20%	30%	40%	50%	10%	20%	30%	40%	50%
	1st Movement	0.2015	0.2215	0.2415	0.2615	0.2915	0.202	0.232	0.262	0.292	0.312
	2nd Movement	0.2195	0.2495	0.2795	0.3095	0.3295	0.2202	0.2402	0.2602	0.2802	0.3102
	3rd Movement	0.184	0.202	0.242	0.262	0.282	0.1839	0.2139	0.2339	0.2639	0.2939
	4th Movement	0.1904	0.2204	0.2404	0.2604	0.2884	0.1893	0.2093	0.2393	0.2693	0.2893

Table 4.30: Performance of CPD on iEMG data of amputee subjects to recover structured missing data of various percentages for day one

Day 1											
Mean of 2 subjects		1st half					2nd half				
		10 %	20 %	30 %	40 %	50 %	10 %	20 %	30 %	40 %	50 %
	1st Movement	0.1	0.12	0.14	0.16	0.19	0.0995	0.1295	0.1595	0.1895	0.2095
	2nd Movement	0.0985	0.1285	0.1585	0.1885	0.2085	0.1005	0.1205	0.1505	0.1705	0.2005
	3rd Movement	0.098	0.116	0.156	0.176	0.196	0.0995	0.1295	0.1495	0.1795	0.2095
	4th Movement	0.1015	0.1315	0.1515	0.1715	0.1995	0.0995	0.1195	0.1495	0.1795	0.1995

Table 4.31: Performance of CPD on iEMG data of amputee subjects to recover structured missing data of various percentages for day two

Day 2											
Mean of 2 subjects		1st half					2nd half				
		10 %	20 %	30 %	40 %	50 %	10 %	20 %	30 %	40 %	50 %
	1st Movement	0.117	0.147	0.177	0.207	0.227	0.1165	0.1365	0.1565	0.1765	0.2065
	2nd Movement	0.1185	0.1385	0.1585	0.1785	0.2085	0.1205	0.1505	0.1805	0.2105	0.2305
	3rd Movement	0.112	0.142	0.162	0.192	0.222	0.1135	0.1315	0.1715	0.1915	0.2115
	4th Movement	0.1165	0.1365	0.1665	0.1965	0.2165	0.1145	0.1445	0.1645	0.1845	0.2125

Table 4.32: Performance of CPD on iEMG data of amputee subjects to recover structured missing data of various percentages for day three

Day 3											
Mean of 2 subjects		1st half					2nd half				
		10 %	20 %	30 %	40 %	50 %	10 %	20 %	30 %	40 %	50 %
	1st Movement	0.134	0.154	0.174	0.194	0.224	0.1335	0.1635	0.1935	0.2235	0.2435
	2nd Movement	0.1385	0.1685	0.1985	0.2285	0.2485	0.1405	0.1605	0.1805	0.2005	0.2305
	3rd Movement	0.126	0.144	0.184	0.204	0.224	0.1275	0.1575	0.1775	0.2075	0.2375
	4th Movement	0.1315	0.1615	0.1815	0.2015	0.2295	0.1295	0.1495	0.1795	0.2095	0.2295

Table 4.33: Performance of CPD on iEMG data of amputee subjects to recover structured missing data of various percentages for day four

Day 4											
Mean of 2 subjects		1st half					2nd half				
		10 %	20 %	30 %	40 %	50 %	10 %	20 %	30 %	40 %	50 %
	1st Movement	0.151	0.181	0.211	0.241	0.261	0.1505	0.1705	0.1905	0.2105	0.2405
	2nd Movement	0.1585	0.1785	0.1985	0.2185	0.2485	0.1605	0.1905	0.2205	0.2505	0.2705
	3rd Movement	0.14	0.17	0.19	0.22	0.25	0.1415	0.1595	0.1995	0.2195	0.2395
	4th Movement	0.1465	0.1665	0.1965	0.2265	0.2465	0.1445	0.1745	0.1945	0.2145	0.2425

Table 4.34: Performance of CPD on iEMG data of amputee subjects to recover structured missing data of various percentages for day five

Day 5											
Mean of 2 subjects		1st half					2nd half				
		10 %	20 %	30 %	40 %	50 %	10 %	20 %	30 %	40 %	50 %
	1st Movement	0.168	0.188	0.208	0.228	0.258	0.1675	0.1975	0.2275	0.2575	0.2775
	2nd Movement	0.1785	0.2085	0.2385	0.2685	0.2885	0.1805	0.2005	0.2205	0.2405	0.2705
	3rd Movement	0.154	0.172	0.212	0.232	0.252	0.1555	0.1855	0.2055	0.2355	0.2655
	4th Movement	0.1615	0.1915	0.2115	0.2315	0.2595	0.1595	0.1795	0.2095	0.2395	0.2595

Table 4.35: Performance of CPD on iEMG data of amputee subjects to recover structured missing data of various percentages for day six

Day 6											
Mean of 2 subjects		1st half					2nd half				
		10 %	20 %	30 %	40 %	50 %	10 %	20 %	30 %	40 %	50 %
	1st Movement	0.185	0.215	0.245	0.275	0.295	0.1845	0.2045	0.2245	0.2445	0.2745
	2nd Movement	0.1985	0.2185	0.2385	0.2585	0.2885	0.2005	0.2305	0.2605	0.2905	0.3105
	3rd Movement	0.168	0.198	0.218	0.248	0.278	0.1695	0.1875	0.2275	0.2475	0.2675
	4th Movement	0.1765	0.1965	0.2265	0.2565	0.2765	0.1745	0.2045	0.2245	0.2445	0.2725

Table 4.36: Performance of CPD on iEMG data of amputee subjects to recover structured missing data of various percentages for day seven

Day 7											
Mean of 2 subjects		1st half					2nd half				
		10 %	20 %	30 %	40 %	50 %	10 %	20 %	30 %	40 %	50 %
	1st Movement	0.202	0.222	0.242	0.262	0.292	0.2015	0.2315	0.2615	0.2915	0.3115
	2nd Movement	0.2185	0.2485	0.2785	0.3085	0.3285	0.2205	0.2405	0.2605	0.2805	0.3105
	3rd Movement	0.182	0.2	0.24	0.26	0.28	0.1835	0.2135	0.2335	0.2635	0.2935
	4th Movement	0.1915	0.2215	0.2415	0.2615	0.2895	0.1895	0.2095	0.2395	0.2695	0.2895

4.4.3 PERFORMANCE OF CP-WOPT

Fig. 4.8(c) and Fig. 4.9(c) show the performance of our proposed framework CP-WOPT. It can be seen for both halves of Day 1 that as the percentage of missing data increases from 10% to 50%, the RME increases slightly from 0.083 to 0.087. Likewise, for Day 1 to 7, RME increases slightly from 0.080 to 0.099. Fig. 4.8(c) and Fig. 4.9(c) shows that as total amount of observed data increase from day 1 to 7 along with missing data, overall the RME increases just slightly from 0.08 to 0.09 which shows robustness of our framework against large percentage of missing data.

Fig. 4.10 shows original iEMG data with all movements hidden in noise and then the signal obtained after applying CP-WOPT to it. Fig. 4.10(a) shows the original noisy iEMG signal comprising all four movements hidden in noise, which makes classification of movements difficult. Fig. 4.10(b) shows the same iEMG signal with CP-WOPT applied on it which recovers the missing movement epochs efficiently from noisy iEMG signal. If we compare both iEMG signals, it can be seen that Fig. 4.10(b) is much more suitable for classification as compared to Fig. 4.10(b). Performance of CP-WOPT to recover structured missing data of various percentages for day one to seven on iEMG data of healthy subjects is presented in Tables 4.37 - 4.43 respectively. Performance of CPD to recover structured missing data of various percentages for day one to seven on iEMG data of amputee subjects has been presented from Table 4.44 to Table 4.50 respectively.

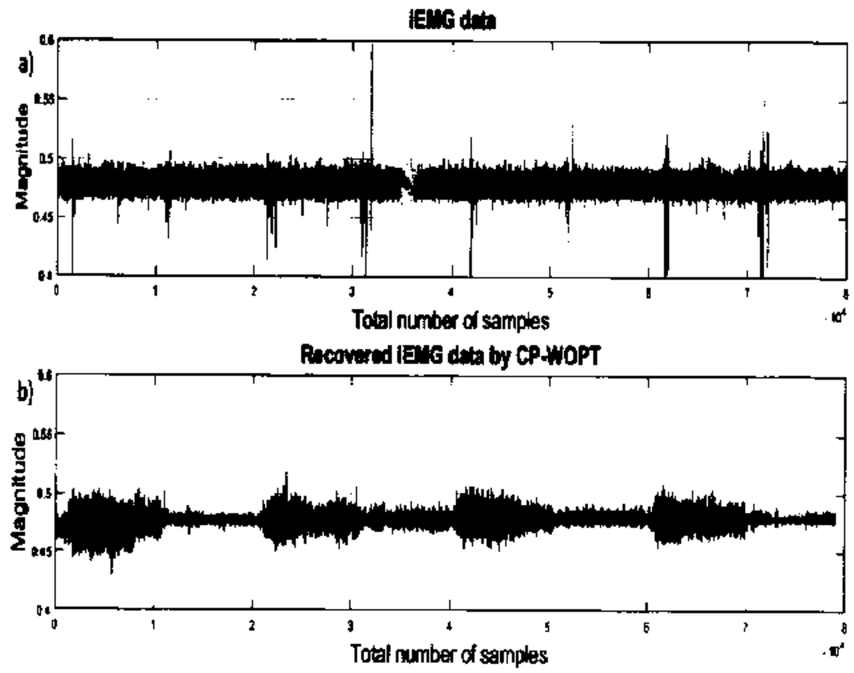

Fig. 4.10 a) Original iEMG data with movements hidden in noise (b) Recovered movements by CP-WOPT

Table 4.37: Performance of CP-WOPT on iEMG data of healthy subjects to recover structured missing data of various percentages for day one

Day 1											
Mean of 10 subject		1st half					2nd half				
		10%	20%	30%	40%	50%	10%	20%	30%	40%	50%
	1st Movement	0.08036	0.08136	0.08336	0.08636	0.08836	0.08033	0.08133	0.08233	0.08433	0.08633
	2nd Movement	0.08033	0.08233	0.08353	0.08453	0.08633	0.08036	0.08236	0.08336	0.08536	0.08636
	3rd Movement	0.08033	0.08235	0.08435	0.08635	0.08735	0.08035	0.08135	0.08335	0.08535	0.08635
	4th Movement	0.08036	0.08135	0.0808	0.0805	0.0825	0.08036	0.08336	0.08536	0.08636	0.08836

Table 4.38: Performance of CP-WOPT on iEMG data of healthy subjects to recover structured missing data of various percentages for day two

Day 2											
Mean of 10 subject		1st half					2nd half				
		10%	20%	30%	40%	50%	10%	20%	30%	40%	50%
	1st Movement	0.08236	0.08336	0.08436	0.08636	0.08836	0.08233	0.08333	0.08533	0.08833	0.09033
	2nd Movement	0.08133	0.08333	0.08433	0.08633	0.08733	0.08146	0.08336	0.08438	0.08536	0.08736
	3rd Movement	0.08235	0.08335	0.08545	0.08735	0.08835	0.08236	0.08435	0.08635	0.08835	0.08955
	4th Movement	0.08135	0.08435	0.08635	0.08735	0.08935	0.08136	0.08236	0.0807	0.0805	0.0826

Table 4.39: Performance of CP-WOPT on iFMG data of healthy subjects to recover structured missing data of various percentages for day three

Day 3											
Mean of 10 subject		1st half					2nd half				
		10%	20%	30%	40%	50%	10%	20%	30%	40%	50%
	1st Movement	0.08436	0.08536	0.08736	0.09036	0.09236	0.08433	0.08533	0.08633	0.08833	0.09033
	2nd Movement	0.08233	0.08433	0.08533	0.08633	0.08833	0.08236	0.08436	0.08536	0.08736	0.08836
	3rd Movement	0.08435	0.08635	0.08835	0.09035	0.09135	0.08435	0.08535	0.08735	0.08935	0.09035
	4th Movement	0.08235	0.08335	0.0807	0.0805	0.0825	0.08236	0.08536	0.08736	0.08836	0.09036

Table 4.40: Perfomance of CP-WOPT on iEMG data of healthy subjects to recover structured missing data of various percentages for day four

Day 4											
Mean of 10 subject		1st half					2nd half				
		10%	20%	30%	40%	50%	10%	20%	30%	40%	50%
	1st Movement	0.08636	0.08736	0.08836	0.09036	0.09236	0.08633	0.08733	0.08933	0.09233	0.09433
	2nd Movement	0.08333	0.08533	0.08633	0.08833	0.08933	0.08336	0.08536	0.08636	0.08736	0.08936
	3rd Movement	0.08635	0.08735	0.08935	0.09135	0.09235	0.08635	0.08835	0.09035	0.09235	0.09335
	4th Movement	0.08335	0.08635	0.08835	0.08935	0.09135	0.08336	0.08436	0.0807	0.0805	0.0825

Table 4.41: Performance of CP-WOPT on iEMG data of healthy subjects to recover structured missing data of various percentages for day five

Day 5											
Mean of 10 subject		1st half					2nd half				
		10%	20%	30%	40%	50%	10%	20%	30%	40%	50%
	1st Movement	0.08836	0.08936	0.09136	0.09436	0.09636	0.08833	0.08933	0.09033	0.09233	0.09433
	2nd Movement	0.08433	0.08633	0.08733	0.08833	0.09033	0.08436	0.08636	0.08736	0.08936	0.09036
	3rd Movement	0.08835	0.09035	0.09235	0.09435	0.09535	0.08835	0.08935	0.09135	0.09335	0.09435
	4th Movement	0.08435	0.08535	0.0807	0.0805	0.0825	0.08436	0.08736	0.08936	0.0[illegible]036	0.09236

Table 4.42: Performance of CP-WOPT on iEMG data of healthy subjects to recover structured missing data of various percentages for day six

Day 6											
Mean of 10 subject		1st half					2nd half				
		10%	20%	30%	40%	50%	10%	20%	30%	40%	50%
	1st Movement	0.09036	0.09136	0.09236	0.09436	0.09636	0.09033	0.09133	0.09333	0.09633	0.09833
	2nd Movement	0.08533	0.08733	0.08833	0.09033	0.09133	0.08536	0.08736	0.08836	0.08936	0.09136
	3rd Movement	0.09035	0.09135	0.09335	0.09535	0.09635	0.09035	0.09235	0.09435	0.09635	0.09735
	4th Movement	0.08535	0.08835	0.09035	0.09135	0.09335	0.08536	0.08636	0.0807	0.0805	0.0825

Table 4.43: Performance of CP-WOPT on iEMG data of healthy subjects to recover structured missing data of various percentages for day seven

Day 7											
Mean of 10 subject		1st half					2nd half				
		10%	20%	30%	40%	50%	10%	20%	30%	40%	50%
	1st Movement	0.09236	0.09336	0.09536	0.09836	0.10036	0.09233	0.09333	0.09433	0.09633	0.09833
	2nd Movement	0.08633	0.08833	0.08933	0.09033	0.09233	0.08636	0.08836	0.08936	0.09136	0.09236
	3rd Movement	0.09235	0.09435	0.09635	0.09835	0.09935	0.09235	0.09335	0.09535	0.09735	0.09835
	4th Movement	0.08635	0.08735	0.0807	0.0805	0.0825	0.08636	0.08936	0.09136	0.09236	0.09436

Table 4.44: Performance of CP-WOPT on iEMG data of amputee subjects to recover structured missing data of various percentages for day one

Day 1											
Mean of 10 subject		1st half					2nd half				
		10%	20%	30%	40%	50%	10%	20%	30%	40%	50%
	1st Movement	0.0803	0.0813	0.0833	0.0863	0.0883	0.0803	0.0813	0.0823	0.0843	0.0863
	2nd Movement	0.08035	0.08235	0.08335	0.08435	0.08635	0.0803	0.0823	0.0833	0.0853	0.0863
	3rd Movement	0.0803	0.0823	0.0843	0.0863	0.0873	0.0803	0.0813	0.0833	0.0853	0.0863
	4th Movement	0.08035	0.08135	0.0807	0.0805	0.0825	0.08035	0.08335	0.08535	0.08635	0.08835

Table 4.45: Performance of CP-WOPT on iEMG data of amputee subjects to recover structured missing data of various percentages for day two

Day 2											
Mean of 10 subject		1st half					2nd half				
		10%	20%	30%	40%	50%	10%	20%	30 %	40 %	50 %
	1st Movement	0.0823	0.0833	0.0843	0.0863	0.0883	0.0823	0.0833	0.0853	0.0883	0.0903
	2nd Movement	0.08135	0.08335	0.08435	0.08635	0.08735	0.0813	0.0833	0.0843	0.0853	0.0873
	3rd Movement	0.0823	0.0833	0.0853	0.0873	0.0883	0.0823	0.0843	0.0863	0.0883	0.0893
	4th Movement	0.08135	0.08435	0.08635	0.08735	0.08935	0.08135	0.08235	0.0807	0.0805	0.083

Table 4.46: Performance of CP-WOPT on iEMG data of amputee subjects to recover structured missing data of various percentages for day three

Day 3											
Mean of 10 subject		1st half					2nd half				
		10%	20%	30%	40%	50%	10%	20%	30%	40%	50%
	1st Movement	0.0843	0.0853	0.0873	0.0903	0.0923	0.0843	0.0853	0.0863	0.0883	0.0903
	2nd Movement	0.08235	0.08435	0.08535	0.08635	0.08835	0.0823	0.0843	0.0853	0.0873	0.0883
	3rd Movement	0.0843	0.0863	0.0883	0.0903	0.0913	0.0843	0.0853	0.0873	0.0893	0.0903
	4th Movement	0.08235	0.08335	0.0807	0.0805	0.0825	0.08235	0.08535	0.08735	0.08835	0.09035

Table 4.47: Performance of CP-WOPT on iEMG data of amputee subjects to recover structured missing data of various percentages for day four

Day 4											
Mean of 10 subject		1st half					2nd half				
		10%	20%	30%	40%	50%	10%	20%	30%	40%	50%
	1st Movement	0.0863	0.0873	0.0883	0.0903	0.0923	0.0863	0.0873	0.0893	0.0923	0.0943
	2nd Movement	0.08335	0.08535	0.08635	0.08835	0.08935	0.0833	0.0853	0.0863	0.0873	0.0893
	3rd Movement	0.0863	0.0873	0.0893	0.0913	0.0923	0.0863	0.0883	0.0903	0.0923	0.0933
	4th Movement	0.08335	0.08635	0.08835	0.08935	0.09135	0.08335	0.08435	0.0807	0.0805	0.0825

Table 4.48: Performance of CP-WOPT on iEMG data of amputee subjects to recover structured missing data of various percentages for day five

Day 5											
Mean of 10 subject		1st half					2nd half				
		10%	20%	30%	40%	50%	10%	20%	30%	40%	50%
	1st Movement	0.0883	0.0893	0.0913	0.0943	0.0963	0.0883	0.0893	0.0903	0.0923	0.0943
	2nd Movement	0.08435	0.08635	0.08735	0.08835	0.09035	0.0843	0.0863	0.0873	0.0893	0.0903
	3rd Movement	0.0883	0.0903	0.0923	0.0943	0.0953	0.0883	0.0893	0.0913	0.0933	0.0943
	4th Movement	0.08435	0.08535	0.0807	0.0805	0.0825	0.08435	0.08735	0.08935	0.09035	0.09235

Table 4.49: Performance of CP-WOPT on iEMG data of amputee subjects to recover structured missing data of various percentages for day six

Day 6											
Mean of 10 subject		1st half					2nd half				
		10%	20%	30%	40%	50%	10%	20%	30 %	40 %	50 %
	1st Movement	0.0903	0.0913	0.0923	0.0943	0.0963	0.0903	0.0913	0.0933	0.0963	0.0983
	2nd Movement	0.08535	0.08735	0.08835	0.09035	0.09135	0.0853	0.0873	0.0883	0.0893	0.0913
	3rd Movement	0.0903	0.0913	0.0933	0.0953	0.0963	0.0903	0.0923	0.0943	0.0963	0.0973
	4th Movement	0.08535	0.08835	0.09035	0.09135	0.09335	0.08535	0.08635	0.0807	0.0805	0.0825

Table 4.50: Performance of CP-WOPT on iEMG data of amputee subjects to recover structured missing data of various percentages for day seven

Day 7											
Mean of 10 subject		1st half					2nd half				
		10%	20%	30%	40%	50%	10%	20%	30%	40%	50%
	1st Movement	0.0923	0.0933	0.0953	0.0983	0.1003	0.0923	0.0933	0.0943	0.0963	0.0983
	2nd Movement	0.08635	0.08835	0.08935	0.09035	0.09235	0.0863	0.0883	0.0893	0.0913	0.0923
	3rd Movement	0.0923	0.0943	0.0963	0.0983	0.0993	0.0923	0.0933	0.0953	0.0973	0.0983
	4th Movement	0.08635	0.08735	0.0807	0.0805	0.0825	0.08635	0.08935	0.09135	0.09235	0.09435

4.5 Benchmarking comparison

Unstructured and structured missing data has been recovered in simulated, sEMG and iEMG data by employing NMF, CP and CP-WOPT. Results show that overall performance of NMF and CP/PARAFAC on EMG data was between 0.3 and 0.6 (in terms of RME) whereas CP-WOPT outperformed both NMF and CP/PARAFAC as it recovered missing data with much more accuracy of up to 0.1 i.e. in extreme case when size of missing data was too large.

4.6 Research contribution

Research contributions of first study are:

a) For the first time, missing data in EMG signals are recovered using the tensor factorization-based method.

b) We compare both matrix factorization, and tensor factorization-based approaches to recover missing data in noisy simulated data and real-world EMG data to show that the tensor-based approach outperforms matrix factorization based approach.

c) We address the problem of missing data in extreme cases when up to half *consecutive* EMG samples of a particular channel are missing. Our proposed framework successfully recovers the missing data even in such an extreme case.

Research contributions of second study are:

1) Missing data from iEMG signals of both healthy and amputee subjects are recovered using state-of-the-art tensor factorization methods.

2) We compare the performance of matrix and tensor factorization methods to recover missing data in real-world iEMG signals. Furthermore, we utilize EMG data of seven days and test the performance of both matrix and tensor factorization methods in a large multiday dataset.

3) We consider the case when up to 50% iEMG data is missing from day 1 to 7 and test the performance of both matrix and tensor factorization methods. We show that CP-WOPT outperformed both NMF and CPD to recover missing data even in the worst-case scenario.

4.7 Summary

Both matrix (NMF) and tensor factorization methods (CPD and CP-WOPT) are employed on sEMG of healthy subjects and iEMG signals of healthy as well as amputee subjects to explore their ability to recover missing data. To further explore the effect of the size of missing data on performance of factorization methods, we gradually increased the total size of missing iEMG data by including data from day 1 to day 7. For each day we increased the percentages of missing data from 10% to 50%. We removed data from two halves. From first half we removed data from 10% to 50% i.e. 10% means removal of data from initial 10% values. Removal of 10% from second half means removing last 10% values. The main finding is that as the size of missing data increases, the performance of NMF and CPD degrades substantially, however, CP-WOPT outperformed both NMF and CPD in terms of RME.

The results of different analyses show that CP-WOPT not only outperformed both NMF and CPD, but it also performed much better over sEMG data of single day and iEMG data of multiple days, and hence, the results were consistent with the notion that CP-WOPT performs comparatively better when the size of missing data is too large [33, 54].

Chapter 5.

Conclusion and Further Work

Two different studies were performed in this dissertation. The main objectives of the first study was to evaluate the performance of weighted CP in comparison with classical matrix and tensor factorization methods on synthetic and healthy sEMG data of single day. The objective of the second study were to test the performance of weighted CP on intramuscular EMG data recorded over multiples days using both healthy and amputee subjects.

This chapter discusses the outcomes and main findings after evaluating all these objectives.

5.1 Conclusion

In this dissertation, mainly matrix and tensor factorization algorithms are explored for engineering applications such as possible improvement in myoelectric control schemes (prosthetic arm) and evaluate the long-term performance with both the surface and intramuscular EMG data. Performance comparison of NMF, PARAFAC with CP-WOPT showed that CP-WOPT outperformed both NMF and PARAFAC in sEMG and iEMG signals of both healthy and amputee subjects for single as well as multiday data. CP-WOPT showed significant performance in recovering both unstructured and structured missing data.

Proposed framework has been tested with Non-Negative Matrix Factorization (NMF) and Parallel Factor Analysis (PARAFAC) on simulated as well as on offline EMG signals having unstructured missing values (randomly missing data ranging from 60% to 95%) and structured missing values. In the case of structured missing data having different channels, the percentage of missing data of a channel goes up to 50% for different movements. It has been observed empirically that the proposed framework recovers the missing data with relatively much improved accuracy in terms of Relative Mean Error (up to 50% and 30 % for unstructured and structured missing data respectively) as compared to matrix factorization methods even when the portion of unstructured and structured missing data reaches up to 95% and 50%, respectively. Moreover results of our proposed framework are much better than benchmark paper 1 and 2, in terms of recovering missing data. Proposed frameworks are extensively tested on moderate to large sEMG and iEMG datasets for both healthy and amputee subjects which were missing in benchmark paper 1 and 2.

Novel tensor factorization framework based on CP-WOPT has been designed based on strong mathematical foundations of multilinear algebra. Moreover, proposed tensor factorization framework has been applied to real-life surface and intramuscular EMG data of both healthy and amputee subjects.

Classification performed using test with recovered samples is expected to yield far better accuracy resulting in improved myoelectric control.

5.2 Further work

Although findings of chapter 4 are promising, but the analysis were performed offline and hence the generalization of CP-WOPT in real time needs to be investigated in future. Limited number of amputee subjects also limits the generalization of results. Moreover, database for this study included data of only seven days. Therefore, future studies will include more number of both able-bodied and amputee subjects and protocol will be made to include data collection over weeks.

Preliminary study based on exploring strength of CP-WOPT to recover missing data proved that CP-WOPT could improve the field of myoelectric control by improving classification accuracy. Again, this study was also performed offline and with only ten able-bodied subjects which limits the generalization of results. Therefore, future studies will include data spanning over a longer duration of time, for example from 2 to 4 weeks. CP-WOPT will also be explored in real-time application. Furthermore, performance of classification before and after recovering missing data will be compared in future study.

BIBLIOGRAPHY

[1] M. Zia ur Rehman, A. Waris, S. Gilani, M. Jochumsen, I. Niazi, M. Jamil, *et al.*, "Multiday EMG-based classification of hand motions with deep learning techniques," *Sensors*, vol. 18, p. 2497, 2018.

[2] E. Acar, D. M. Dunlavy, T. G. Kolda, and M. Mørup, "Scalable tensor factorizations for incomplete data," *Chemometrics and Intelligent Laboratory Systems*, vol. 106, pp. 41-56, 2011.

[3] M. Zia ur Rehman, S. Gilani, A. Waris, I. Niazi, G. Slabaugh, D. Farina, *et al.*, "Stacked sparse autoencoders for EMG-based classification of hand motions: A comparative multi day analyses between surface and intramuscular EMG," *Applied Sciences*, vol. 8, p. 1126, 2018.

[4] E. Scheme and K. Englehart, "Electromyogram pattern recognition for control of powered upper-limb prostheses: state of the art and challenges for clinical use," *Journal of Rehabilitation Research & Development*, vol. 48, 2011.

[5] A. Phinyomark, P. Phukpattaranont, and C. Limsakul, "A review of control methods for electric power wheelchairs based on electromyography signals with special emphasis on pattern recognition," *IETE Technical Review*, vol. 28, pp. 316-326, 2011.

[6] T. S. Saponas, D. S. Tan, D. Morris, R. Balakrishnan, J. Turner, and J. A. Landay, "Enabling always-available input with muscle-computer interfaces," in *Proceedings of the 22nd annual ACM symposium on User interface software and technology*, 2009, pp. 167-176.

[7] A. NABER, "Stationary Wavelet Processing and Data Imputing in Myoelectric Pattern Recognition on an Embedded System."

[8] D. Farina, N. Jiang, H. Rehbaum, A. Holobar, B. Graimann, H. Dietl, *et al.*, "The extraction of neural information from the surface EMG for the control of upper-limb prostheses: emerging avenues and challenges," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 22, pp. 797-809, 2014.

[9] C. Setz, J. Schumm, C. Lorenz, B. Arnrich, and G. Tröster, "Using ensemble classifier systems for handling missing data in emotion recognition from physiology: one step towards a practical system," in *2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops*, 2009, pp. 1-8.

[10] A. Waris, I. K. Niazi, M. Jamil, O. Gilani, K. Englehart, W. Jensen, *et al.*, "The effect of time on EMG classification of hand motions in able-bodied and transradial amputees," *Journal of Electromyography and Kinesiology*, vol. 40, pp. 72-80, 2018.

[11] J. Yousefi and A. Hamilton-Wright, "Characterizing EMG data using machine-learning tools," *Computers in biology and medicine*, vol. 51, pp. 1-13, 2014.

[12] N. Jiang, K. B. Englehart, and P. A. Parker, "Extracting simultaneous and proportional neural control information for multiple-DOF prostheses from the surface electromyographic signal," *IEEE transactions on Biomedical Engineering*, vol. 56, pp. 1070-1080, 2008.

35th annual international conference of the IEEE engineering in medicine and biology society (EMBC), 2013, pp. 4223-4226.

[28] G. T. YEE, "Missing Data Problem in Random Electrocardiogram Signal Processing," Universiti Teknologi Malaysia, 2014.

[29] S.-H. Liew, Y.-H. Choo, and Y. F. Low, "Missing Values Imputation using Similarity Matching Method for Brainprint Authentication," *INTERNATIONAL JOURNAL OF ADVANCED COMPUTER SCIENCE AND APPLICATIONS*, vol. 9, pp. 364-370, 2018.

[30] S.-H. Kim, H.-J. Yang, S.-H. Kim, and G.-S. Lee, "PhysioCover: Recovering the Missing Values in Physiological Data of Intensive Care Units," *International Journal of Contents*, vol. 10, 2014.

[31] Q. Ding, J. Han, X. Zhao, and Y. Chen, "Missing-data classification with the extended full-dimensional Gaussian mixture model: Applications to EMG-based motion recognition," *IEEE Transactions on Industrial Electronics*, vol. 62, pp. 4994-5005, 2015.

[32] A. K. A. Al-naqeeb, A. A. Ibrahim, and Q. S. Tawfeeq, "New Nonlinearities Interpolation Approach Applied to Surface EMG Signal," in *BIOSIGNALS*, 2014, pp. 171-177.

[33] M. Akmal, S. Zubair, M. Jochumsen, E. N. Kamavuako, and I. K. Niazi, "A tensor-based method for completion of missing electromyography data," *Ieee Access*, vol. 7, pp. 104710-104720, 2019.

[34] T. Tsuji, O. Fukuda, M. Murakami, and M. Kaneko, "An EMG controlled pointing device using a neural network," *Transactions of the Society of Instrument and Control Engineers*, vol. 37, pp. 425-431, 2001.

[35] O. Fukuda, T. Tsuji, and M. Kaneko, "An EMG controlled pointing device using a neural network," in *IEEE SMC'99 Conference Proceedings. 1999 IEEE International Conference on Systems, Man, and Cybernetics (Cat. No. 99CH37028)*, 1999, pp. 63-68.

[36] O. Fukuda, J. Arita, and T. Tsuji, "An EMG-controlled omnidirectional pointing device using a HMM-based neural network," in *Proceedings of the International Joint Conference on Neural Networks, 2003.*, 2003, pp. 3195-3200.

[37] N. Bu, T. Hamamoto, T. Tsuji, and O. Fukuda, "FPGA implementation of a probabilistic neural network for a bioelectric human interface," in *The 2004 47th Midwest Symposium on Circuits and Systems, 2004. MWSCAS'04.*, 2004, pp. iii-29.

[38] J. Kim, S. Mastnik, and E. André, "EMG-based hand gesture recognition for realtime biosignal interfacing," in *Proceedings of the 13th international conference on Intelligent user interfaces*, 2008, pp. 30-39.

[39] A. D. Chan and G. C. Green, "Myoelectric control development toolbox," *CMBES Proceedings*, vol. 30, 2007.

[40] W. Putnam and R. B. Knapp, "Real-time computer control using pattern recognition of the electromyogram," in *Proceedings of the 15th Annual International Conference of the IEEE Engineering in Medicine and Biology Societ*, 1993, pp. 1236-1237.

[41] G. Tsenov, A. Zeghbib, F. Palis, N. Shoylev, and V. Mladenov, "Neural networks for online classification of hand and finger movements using surface EMG signals,"

in *2006 8th Seminar on Neural Network Applications in Electrical Engineering*, 2006, pp. 167-171.

[42] R. Rosenberg, "The biofeedback pointer: EMG control of a two dimensional pointer," in *Digest of Papers. Second International Symposium on Wearable Computers (Cat. No. 98EX215)*, 1998, pp. 162-163.

[43] K. K. Jung, J. W. Kim, H. K. Lee, S. B. Chung, and K. H. Eom, "EMG pattern classification using spectral estimation and neural network," in *SICE Annual Conference 2007*, 2007, pp. 1108-1111.

[44] E. M. El-Daydamony, M. El-Gayar, and F. Abou-Chadi, "A computerized system for semg signals analysis and classifieation," in *2008 National Radio Science Conference*, 2008, pp. 1-7.

[45] F. Mobasser and K. Hashtrudi-Zaad, "A method for online estimation of human arm dynamics," in *2006 International Conference of the IEEE Engineering in Medicine and Biology Society*, 2006, pp. 2412-2416.

[46] P. Geethanjali and K. Ray, "Identification of motion from multi-channel EMG signals for control of prosthetic hand," *Australasian physical & engineering sciences in medicine*, vol. 34, pp. 419-427, 2011.

[47] A. Waris and E. N. Kamavuako, "Effect of threshold values on the combination of emg time domain features: Surface versus intramuscular emg," *Biomedical Signal Processing and Control*, vol. 45, pp. 267-273, 2018.

[48] X. Chen, X. Zhang, Z.-Y. Zhao, J.-H. Yang, V. Lantz, and K.-Q. Wang, "Hand gesture recognition research based on surface EMG sensors and 2D-accelerometers," in *2007 11th IEEE International Symposium on Wearable Computers*, 2007, pp. 11-14.

[49] A. Alkan and M. Günay, "Identification of EMG signals using discriminant analysis and SVM classifier," *Expert Systems with Applications*, vol. 39, pp. 44-47, 2012.

[50] M. Jochumsen, A. Waris, and E. N. Kamavuako, "The effect of arm position on classification of hand gestures with intramuscular emg," *Biomedical Signal Processing and Control*, vol. 43, pp. 1-8, 2018.

[51] S. Keleş and A. Subaşı, "Classification of EMG signals using decision tree methods," in *Third International Symposium on Sustainable Development (ISSD'12)*, 2012, p. 354.

[52] G. Cui, L. Gui, Q. Zhao, A. Cichocki, and J. Cao, "Bayesian CP factorization of incomplete tensor for EEG signal application," in *2016 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, 2016, pp. 2170-2173.

[53] A. Farhangfar, L. Kurgan, and J. Dy, "Impact of imputation of missing values on classification error for discrete data," *Pattern Recognition*, vol. 41, pp. 3692-3705, 2008.

[54] E. Acuna and C. Rodriguez, "The treatment of missing values and its effect on classifier accuracy," in *Classification, clustering, and data mining applications*, ed: Springer, 2004, pp. 639-647.

[55] T. Sidekerskiene and R. Damasevicius, "Reconstruction of missing data in synthetic time series using EMD," in *CEUR Workshop Proceedings*, 2016, pp. 7-12.

[56] T. N. Thieu, H.-J. Yang, T. D. Vu, and S.-H. Kim, "Tensor Completion Based on Tucker Factorization for Missing Value Imputation," *International Journal of Bioscience, Biochemistry and Bioinformatics,* vol. 6, p. 34, 2016.

[57] J. Solé-Casals, C. F. Caiafa, Q. Zhao, and A. Cichocki, "Brain-computer interface with corrupted EEG data: A tensor completion approach," *Cognitive Computation,* vol. 10, pp. 1062-1074, 2018.

[58] E. Acar and B. Yener, "Unsupervised multiway data analysis: A literature survey," *IEEE transactions on knowledge and data engineering,* vol. 21, pp. 6-20, 2008.

[59] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM review,* vol. 51, pp. 455-500, 2009.

[60] P. Comon, X. Luciani, and A. L. De Almeida, "Tensor decompositions, alternating least squares and other tales," *Journal of Chemometrics: A Journal of the Chemometrics Society,* vol. 23, pp. 393-405, 2009.

[61] J. Nocedal and S. Wright, *Numerical optimization*: Springer Science & Business Media, 2006.

[62] J. B. Kruskal, "Three-way arrays: rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics," *Linear algebra and its applications,* vol. 18, pp. 95-138, 1977.

[63] M. Mørup, "Applications of tensor (multiway array) factorizations and decompositions in data mining," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery,* vol. 1, pp. 24-40, 2011.

[64] J. D. Carroll and J.-J. Chang, "Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart-Young" decomposition," *Psychometrika,* vol. 35, pp. 283-319, 1970.