POLITECNICO DI MILANO
Facoltà di Ingegneria dei Sistemi

Corso di Laurea in Ingegneria Biomedica

Time-frequency analysis of the ballistocardiogram for sleep staging

Relatore: Prof. Sergio Cerutti
Correlatori: Prof.ssa Anna Maria Bianchi
Prof. Martin Mendez

Elaborato finale di:

Matteo MIGLIORINI 724760
Domenico NISTICO' 721370

Anno Accademico 2009/2010

Sommario

Obiettivi

L’analisi effettuata grazie all’elettroencefalografia (EEG) è una tecnica molto usata per studiare l’attività del sistema nervoso centrale (CNS). Fornisce informazioni legate all’attività cerebrale tramite misure elettriche sullo scalpo dei soggetti. Grazie all’EEG vengono studiati i trattamenti efficaci e la diagnosi del paziente per alcune patologie. Le informazioni che vengono ricavate sono principalmente legate all’analisi spettrale. Lo studio dell’EEG di un paziente dormiente è una branca della medicina molto importante proprio per le sue applicazioni cliniche. Il gold standard per la stadiazione del sonno sui pazienti è basato sull’analisi delle onde cerebrali descritto da Rechtschaffen e Kales. Il metodo più comune è la polisonnografia (PSG) che utilizza, oltre all’EEG, altri segnali fisiologici come l’elettromiogramma (EMG), elettro-oclugramma (EOG), respiro, saturazione di ossigeno nel sangue, elettrocardiogramma (ECG) e analisi visive. La PSG utilizza epoche di 30 secondi per studiare il sonno. Tale metodo è normalmente effettuato in strutture attrezzate e necessita di personale medico per il set up dei sensori, il monitoraggio e l’analisi. Sebbene l’analisi venga effettuata sul computer, è necessario comunque uno specialista del sonno che stadi le epoche. Si può facilmente comprendere lo spreco di risorse sia in termini di costo che di tempo; se poi si aggiunge che l’EEG richiede l’installazione di molti elettrodi sullo scalpo, si può comprendere come il paziente possa non raggiungere una buona qualità del sonno, alterando spesso le diagnosi. Le variazioni nel Sistema Nervoso Autonomo (ANS) durante il passaggio veglia/sonno si riflettono su molti segnali facilmente acquisibili (frequenza cardiaca, pressione, ecc). Questa tesi utilizza il segnale di variabilità cardiaca (HRV) durante il sonno. Esso può essere acquisito in diversi modi, ed è per questo che molti risultati si contraddicono con altri. Infatti l’HRV è suscettibile al rumore rendendo difficili le applicazioni con tale tipo di segnale. Recenti studi hanno affiancato il respiro all’HRV dimostrando l’utilità del segnale

cardiorespiratorio per la discriminazione degli stadi del sonno. Il vantaggio di questo segnale consiste nella facilità di misura e grazie all'impiego di sensori integrati può essere acquisito anche da persone non esperte.

Questa tesi ha l'obiettivo di creare non solo un algoritmo di detezione automatica del sonno, ma anche di valutare alcuni parametri clinici come la *sleep efficiency*, la percentuale di sonno REM rispetto al sonno totale (REM%TST), e la latenza del primo ciclo REM basandosi su parametri spettrali.

Lo scopo è quindi sviluppare un algoritmo di *machine-learning* che partendo dal segnale HRV, dal respiro e dal movimento, riesca a identificare e apprendere le caratteristiche che distinguono le fasi del sonno, stadiando nuove registrazioni.

Materiali e metodi

I dati sono stati acquisiti da 11 adulti (età fra i 20 e i 54 anni) al centro del sonno Finnish Institute of Occupational Health (FIOH). Dopo una registrazione notturna di riferimento, ogni soggetto ha effettuato due registrazioni: una durante la notte, e una durante il giorno dopo aver trascorso la notte a lavorare. I segnali sono stati studiati usando lo standard R&K sui segnali EEG, EOG e EMG. La stadiazione è stata effettuata da medici specializzati. Gli intervalli RR (RRI) sono stati estrapolati dal segnale ECG tramite il software Somnologica. Inoltre il segnale BCG multicanale è stato acquisito tramite sensori interni a un materasso (*bed sensor*), basati su multielettrodi Emfit. Sono stati acquisiti sia l'Heart Beat Interval (HBI), circa l'88% del segnale, e sia il movimento dal *bed sensor*. La macrostruttura del sonno è stata fornita da un esperto medico per ogni soggetto secondo i criteri di R&K.

Estrazione delle features

La prima parte di questo lavoro si focalizza sul *signal processing* e sulla ricerca delle features che caratterizzano gli stadi del sonno, al fine di ottenere una detezione automatica delle epoche. Per ognuna delle 17 registrazioni, l'HRV e il respiro sono stati sincronizzati con il movimento e l'ipnogramma medico, e perciò sono state estratte le features per ognuno dei due segnali. La procedura consiste nel filtrare i due segnali in 3 bande: Very Low Frequencies (VLF), Low Frequencies (LF) e High Frequencies (HF), dopo aver

eliminato eventuali outlier dovuti ad artefatti da movimento con un filtro mediano. Le features sono quindi state normalizzate rispetto alla potenza totale (TP) e sono stati calcolati anche i loro rapporti (LF/VLF, LF/HF e HF/VLF). Infine è stata calcolata anche la cross-correlazione tra i due segnali nelle stesse bande; tale metodo è utile per valutare il grado di somiglianza dei due segnali in ogni singolo stadio. I valori sono stati mediati per ogni epoca, mentre invece sono state calcolate due features su intervalli di tempo più ampi, 5 epoche (150 secondi), come la media e la deviazione standard. Le features scelte sono quelle che massimizzano le performance migliori usando il Discriminante Lineare (LD) o Quadratico (QD).

Il segnale viene elaborato con un filtro Recursive Least Square (RLS) e con la decomposizione tramite Trasformata Wavelet Discreta (WDT).

Non sono state prese in analisi le registrazioni di 5 pazienti a causa di anormalità presenti nel segnale, come una bassa variabilità cardiaca durante le fasi del sonno e la presenza di battiti ectopici.

Questo lavoro vuole mostrare l'importanza dell'informazione che dà il respiro; a tal fine è stata calcolata la performance dei classificatori usando le features estratte solo da HRV, solo da respiro o da entrambi.

Selezione delle features

L'algoritmo di selezione delle features è importante per i sistemi di riconoscimento e classificazione; se si considera uno spazio di features a tante dimensioni, la performance del classificatore peggiorerà in termini sia di costo computazionale che di precisione. Quest'ultima può diminuire a causa della ridondanza delle features, ma anche perché poche features possono attenuare la dimensionalità, quando il training set è limitato, portando all'overtraining. Dall'altra parte, una riduzione del numero di features potrebbe portare a una perdita di discriminazione e perciò a una più bassa accuratezza di riconoscimento del sistema. Il criterio per la scelta del miglior subset di features è basato sul Kappa di Cohen e utilizza un algoritmo di selezione sequenziale *forward* su tutto lo spazio delle features. L'algoritmo inizialmente ha uno spazio vuoto e ad ogni iterazione viene scelta la feature che massimizza il Kappa index, aggiungendola al subset. Il calcolo del Kappa index è basato sul Discriminante Lineare o Quadratico.

Le tre fasi del sonno sono state analizzate separatamente e perciò ognuna ha un suo subset

ottimo.

Classificazione Automatica

La stadiazione del sonno è stata effettuata utilizzando le features scelte precedentemente e la classificazione one-versus-all con differenti metodologie:

- Discriminante Lineare (LD): la classificazione viene effettuata utilizzando una combinazione lineare di features scelte tramite l'algoritmo SFS, basato su LD.
- Discriminante Quadratico (QD): separa i dati di due classi di oggetti o eventi con una superficie quadratica. Le features usate sono selezionate dall'algoritmo SFS, basato sul QD.
- *K*-Nearest-Neighbor (KNN): riconosce pattern per la classificazione di oggetti basandosi sulle caratteristiche degli oggetti vicini a quello considerato. Le features sono selezionate tramite SFS, basato su LD.
- Feed-Forward Neural Network (FFNN): per ogni classe, la rete fornisce una probabilità di appartenere o meno alla classe considerata. La rete è addestrata con le features scelte tramite SFS basato su LD. L'ottimizzazione della rete è stata effettuata facendo variare il numero di neuroni nello strato nascosto e inizializzando i pesi con valori diversi varie volte.

Tutte le tecniche di *machine-learning* sono state implementate usando l'algoritmo Leave-One-Out. In più sono state applicate alcuni espedienti per aumentare le performance del classificatore. È stato utilizzato un filtro mediano (di ordine ottimo calcolato iterativamente) per eliminare alcune brusche variazioni nell'ipnogramma. È stata classificata come WAKE l'epoca in cui era presente una forte attività motoria e, inoltre, se due movimenti sono presenti in un intervallo di tempo ravvicinato, l'intervallo intero viene classificato come fase WAKE. L'intervallo ottimo viene scelto sulla base della massimizzazione del Kappa index medio di tutti i pazienti.

Risultati

I risultati sono stati ottenuti con le features estratte dal modello RLS e dalla decomposizione tramite WDT. Ognuno dei metodi di *signal processing* classifica le fasi del sonno con i propri parametri.

Discriminante Lineare

Il Discriminante Lineare è uno dei classificatori più semplici, con un basso costo computazionale. La tabella 1 mostra i valori medi di 17 registrazione ottenuti tramite analisi basata su RLS e WDT. La miglior classificazione è ottenuta grazie a RLS, raggiungendo un Kappa index di 0.54 e un'accuratezza del 76.39%. La classificazione è ottenuta con le features estratte dal modello RLS e ottiene risultati d'interesse anche nella stima di parametri clinici. L'analisi tramite WDT raggiunge al più gli stessi risultati del RLS; si può notare come il primo ciclo REM, e quindi la sua latenza, sia migliore della classificazione tramite RLS.

	LD classification							
	Acc	**kappa**	**SeAuto**	**SeHyp**	**REM%TST auto**	**REM%TST hyp**	**Lat Auto**	**Lat Hyp**
RLS	76.39 ± 7.61	0.54 ± 0.10	92.33 ± 2.57	85.21 ± 7.25	29.39 ± 15.50	23.42 ± 6.71	87 ± 49	79 ± 33
Wavelet	73.92 ± 15.33	0.51 ± 0.17	92.27 ± 2.48	85.21 ± 7.25	31.31 ± 22.25	23.42 ± 6.71	78±58	79±33

Tabella 1. Media e deviazione standard tra l'accuratezza e indici di somiglianza per le fasi del sonno ottenute tramite RLS e WDT per LD. Acc sta per accuratezza totale, kappa per kappa index, Se è la sleep efficiency, REM%TST è la percentuale di Rem durante il Sonno Totale (TST), e Lat rappresenta la latenza, in minuti, del primo ciclo REM. "Auto" è il suffisso delle precedenti abbreviazioni ed indica i parametri ottenuti tramite detezione automatica, l'altro suffisso "hyp" è legato ai parametri estratti dall'ipnogramma medico.

Discriminante Quadratico

Il discriminante quadratico ha la migliore performance raggiunta rispetto ad ogni classificatore usando le features ottenute tramite RLS sulle 17 registrazioni scelte in precedenza. Raggiunge una classificazione di 0.55 per il kappa index e un'accuratezza del

76.80%. Questo classificatore identifica la latenza del primo ciclo REM con una buona precisione, indicando una qualità del sonno vicino alla predizione medica. La tabella 2 mostra il confronto fra i due tipi di estrazione di features. Le wavelet stimano bene la sleep efficiency, ma non raggiungono buoni risultati con gli altri risultati.

	QD classification							
	Acc	**kappa**	**SeAuto**	**SeHyp**	**REM%TST auto**	**REM%TST hyp**	**Lat Auto**	**Lat Hyp**
RLS	76.81 ± 7.51	0.55 ± 0.10	90.25 ± 4.67	85.21 ± 7.25	30.02 ± 14.49	23.42 ± 6.71	86 ± 52	79 ± 33
Wavelet	69.27 ± 12.64	0.35 ± 0.15	91.65 ± 3.25	85.21 ± 7.25	19.36 ± 25.54	23.42 ± 6.71	104±85	79 ± 33

Tabella 2. Media e deviazione standard tra l'accuratezza e indici di somiglianza per le fasi del sonno ottenute tramite RLS e WDT perQD.

***K*-Nearest-Neighbors**

Il parametro ottimo per questo classificatore è K=25 per le features estratte dal modello RLS, che porta a un Kappa index medio di 0.42 e un'accuratezza del 72%. Si può notare dalla tabella 3 come questo classificatore non abbia performance migliori di altri classificatori nella stadiazione del sonno, soprattutto nell'identificazione dei parametri clinici. La buona classificazione della *sleep efficiency* potrebbe essere considerata casuale basandosi su altri parametri.

	***K*-NN classification**							
	Acc	**kappa**	**SeAuto**	**SeHyp**	**REM%TST auto**	**REM%TST hyp**	**Lat Auto**	**Lat Hyp**
RLS	71.95 ± 7.47	0.42 ± 0.10	78.77 ± 10.06	85.21 ± 7.25	6.80 ± 8.02	23.42 ± 6.71	201±149	79 ± 33
Wavelet	64.75 ± 9.30	0.33 ± 0.10	64.55 ± 14.63	85.21 ± 7.25	0.75 ± 1.10	23.42 ± 6.71	239±122	79 ± 33

Tabella 3. Media e deviazione standard tra l'accuratezza e indici di somiglianza per le fasi del sonno ottenute tramite RLS e WDT per K-NN.

Feed-Forward Neural Network

La tabella 4 mostra i risultati della Feed-Forward Neural Network (FFNN) basata sulla back-propagation. Si può notare come l'accuratezza e il Kappa index siano peggiori di quelli ottenuti con gli altri classificatori, ma anche che l'analisi ottenuta con le features estratte dal RLS presenta una buona predizione della *sleep efficiency*, e quella ottenuta con le features estratte dalle wavelet una buona stima del REM%TST e della latenza del primo ciclo REM. Questi risultati, in ogni caso, non risultano consistenti a causa di un'accuratezza e di un Kappa index troppo bassi. Guardando i singoli ipnogrammi predetti dalla rete neurale si nota che la FFNN non raggiunge una buona predizione. Infatti, molte volte vengono confusi gli stadi WAKE con quelli REM con entrambi i metodi di *signal processing*.

	FFNN classification							
	Acc	**kappa**	**SeAuto**	**SeHyp**	**REM%TST auto**	**REM%TST hyp**	**Lat Auto**	**Lat Hyp**
RLS	67.17 ± 11.88	0.39 ± 0.13	71.60 ± 17.62	85.21 ± 7.25	11.60 ± 14.14	23.42 ± 6.71	200±141	79 ± 33
Wavelet	50.37 ± 18.79	0.24 ± 0.13	48.75 ± 23.47	85.21 ± 7.25	22.08 ± 28.31	23.42 ± 6.71	87±122	79 ± 33

Tabella 4. Media e deviazione standard tra l'accuratezza e indici di somiglianza per le fasi del sonno ottenute tramite RLS e WDT per FFNN.

Conclusioni

Le performance ottenute mostrano che gli ipnogrammi possono essere stimati in un modo sicuro usando i segnali derivanti dal *bed sensor*. LD e QD, combinato con l'estrazione di features tramite modello RLS, possono essere usati per implementare un sistema di classificazione automatica della macrostruttura del sonno da una serie temporale acquisita tramite *bed sensor*. Questi classificatori offrono una buona stima di alcuni indici della qualità del sonno come la *sleep efficiency*, la percentuale di REM durante il sonno totale, e la latenza del primo ciclo REM. I risultati ottenuti dimostrano che l'estrazione tramite RLS è migliore di quella ottenuta tramite decomposizione wavelet. Ci si aspetta risultati migliori di quelli ottenuti tramite *K*-NN e FFNN; ciò forse è dovuto a una selezione delle features basate su LD. Ricerche future possono indagare la possibilità di incrementi di performance di *K*-NN e FFNN con selezione delle features dedicate.

Questi risultati sono in linea con la letteratura. Redmond *et al.* [17] hanno presentato un algoritmo per la detezione automatica delle fasi del sonno dal segnale cardiorespiratorio. L'introduzione dell'attività del movimento, come feature per la stadiazione WAKE, incrementa le performance da k=0.47 dello studio di Redmond *et al.* a k=0.54 di questo studio utilizzando LD.

Abstract

Objectives

Electroencephalogram analysis (EEG) is a very useful technique to investigate the activity of the central nervous system (CNS). It provides information related to the brain activity based on measurements of electrical recordings taken on the scalp of the subjects. Inference and studies about subject's health and effective treatment of some diseases can be carried out analyzing the information obtained from the EEG. Processing of the recorded information mainly concern spectral analysis. Sleep EEG is a very important research branch of medicine, because of the clinical applications. The gold standard method for assessing sleep in humans is the analysis of brain wave patterns (EEG) first described by Rechtschaffen and Kales. The most common sleep analysis method is called polysomnography (PSG), which combines EEG recordings with different physiological signals like electromyography (EMG), electrooculography (EOG), respiratory effort, blood oxygen saturation, electrocardiograms (ECG) and video analysis. In PSG, 30-second epochs of the signals are used for decision making. The method is normally carried out in a controlled hospital environment and needs medical assistance for setting up sensors, monitoring and analysis. Although the analysis is typically computer-assisted, it still requires a sleep expert and is therefore expensive and time consuming. Furthermore, EEG recordings require many electrodes to be glued to the scalp, which makes it very cumbersome and uncomfortable. Changes in the Autonomic Nervous System (ANS) during sleep/wake transitions have been successfully identified as a reliable source of information. Changes in activity of the ANS are reflected in various physiological signals such as heart rate, blood pressure, skin conductance, etc. The main focus of current research is on fluctuations of heart rate variability (HRV) during sleep. However, the way of calculating the HRV is not uniform and therefore results contradict each other. Further, HRV measures are very susceptible to noise. A comfortable application of this technique is

therefore difficult. Recently, several studies have added respiratory signals to the HRV to show the feasibility of using cardiorespiratory signals for discriminating sleep stages. The advantage of cardiorespiratory signals is that they are easy to measure and the sensors can be applied by non-experienced users.

This thesis aims to create an automatic method to detect sleep stages with the goal not only to provide a precise hypnogram, but also to evaluate some clinical parameters such as sleep efficiency, percentage of Total Sleep Time occupied by REM stages (REM%TST), and first REM cycle latency based on the cardiorespiratory spectral parameters.

We attempt to develop a *machine-learning* tool that, starting from some HRV, respiratory and movement traces, can identify and learn from the data itself the distinctive features that characterize the sleep stages, and be then used as a universal tool to recognize the sleep stages in every new trace.

Materials and methods

Sleep recordings were performed on 11 adults (age 20-54 years) at the sleep laboratory of Finnish Institute of Occupational Health (FIOH). Each subject participated with two recordings and these were obtained after baseline night, once during daytime sleep after a night shift of work and once during nighttime sleep. Signals were scored using standard R&K criteria on EEG, EOG and EMG. The sleep scoring, based on standard polysomnographic recordings, was done by expert personnel. R-R intervals (RRI) were computed from the standard ECG signal with the Somnologica software. In addition, the multichannel BCG was recorded with the bed sensor using multiple Emfit electrodes. Both the heart beat interval (HBI), with coverage of 88%, and movement activity were extracted from the bed sensor signals. The macrostructure for each subject were scored by an expert based on Rechtschaffen and Kales' rules.

Feature extraction

The first part of this work focused on processing the signal and searching for features that characterized sleep stages, to be used to develop the automatic detection algorithm. For each of the 17 recordings, the HRV and the respiratory signals has been synchronized with

the movement activity and the medical hypnogram, and afterwards the *features* have been computed for each of the two signals. The procedure consisted in filtering the HRV and the respiratory signal into three bands: Very Low Frequencies (VLF), Low Frequencies (LF) and High Frequencies (HF), after eliminating the outlier values through a optimal median filter. The features have then been normalized with their Total Power (TP) and also the inter-band ratios were used (LF/VLF, LF/HF and HF/VLF). It was also computed the dot-product of HRV and respiratory spectral bands (cross-correlation); this is useful to evaluate their similarity related to sleep stages. The values have been averaged for each epoch, while other features with good information content in distinguishing sleep stages were computed on 5-epochs moving windows (150 seconds): mean and standard deviation. The best combination of features was chosen from which maximizes performances of Linear Discriminant (LD) classifier or Quadratic Discriminant (QD) classifier.

This signal processing was performed by Recursive Least Square (RLS) model and by Wavelet Discrete Transform (WDT) decomposition.

Five patients' data have been eliminated from the analysis because of signal abnormalities, such as lower HRV during sleep stages and the presence of ectopic beats.

This work is oriented in showing the importance of addicted information given by respiratory signal, thus we have evaluated classifiers performances with only HRV features, only respiratory features or both.

Feature Selection

Feature selection algorithms are important to recognition and classification systems because, if a feature space with a large dimension is used, the performance of the classifier will decrease with respect to execution time and to recognition rate. The recognition rate can decrease because of redundant features and of the fact that small number of features can alleviate the course of dimensionality when the training samples set is limited, leading to overtraining. On the other hand, a reduction in the number of features may lead to a loss in the discrimination power and thereby lower the accuracy of the recognition system. In order to determine the best feature subset for Cohen's kappa index criterion, sequential forward feature selection algorithm (SFS) was applied to the complete feature space. The algorithm starts with a null feature set and, for each step, the best feature that maximizes

kappa index value is included with the current feature set. The kappa index computing method is based on the LD or on the QD classifier.
The three sleep stages were analyzed separately, thus there were as many best features subset.

Automatic Classification

The automatic sleep staging was done exploiting the previously evaluated features and one-versus-all method with different methodologies:

- Linear Discriminant (LD): achieves the discrimination by making a classification decision based on the value of a linear combination of features selected by SFS algorithm based on LD.
- Quadratic Discriminant (QD): separates measurements of two classes of objects or events by a quadric surface. Used features were selected by SFS algorithm based on QD.
- *K*-Nearest-Neighbor (KNN): classifies objects based on closest training examples in the feature space. Features were selected by SFS algorithm based on LD.
- Feed-Forward Neural Network (FFNN): for each state, the network provides the probability of being or not the state. The input neurons of the network are the features selected by SFS algorithm based on LD. The optimal networks have been chosen among those obtained with different numbers of hidden layer neurons, and with various restarts.

All the machine-learning techniques have been implemented with the Leave-One-Out (LOO) technique. Some shrewdness has been applied in order to improve algorithm performances. A median filtering (with optimal order computed iteratively) was applied to the obtained hypnogram smoothing peaks. The presence of movement activity forces the classification to WAKE stage and when two moment events occur in a certain interval, all epochs between these are classified as WAKE stages. The research of optimal interval was performed maximizing mean kappa index value on each subject.

Results

The results were obtained from two feature extractor based on RLS model and WDT decomposition. For each signal processing method, sleep stages were classified by their parameters extracted.

Linear Discriminant

The linear discriminant is one of easiest classifier and has low time complexity. Table 1 shows the mean statistical values over the 17 recordings for RLS and wavelet analysis. The best classification is performed by RLS analysis with 0.54 of kappa index and 0.7639 for the accuracy. Classification performed with features extracted with RLS method achieves interesting results even in clinical parameters estimation. Wavelet analysis achieves almost the same results of RLS classification; one can notice that in this case Prediction of the first REM cycle latency is quite better than RLS classification.

	LD classification							
	Acc	**kappa**	**SeAuto**	**SeHyp**	**REM%TST auto**	**REM%TST hyp**	**Lat Auto**	**Lat Hyp**
RLS	76.39 ± 7.61	0.54 ± 0.10	92.33 ± 2.57	85.21 ± 7.25	29.39 ± 15.50	23.42 ± 6.71	87 ± 49	79 ± 33
Wavelet	73.92 ± 15.33	0.51 ± 0.17	92.27 ± 2.48	85.21 ± 7.25	31.31 ± 22.25	23.42 ± 6.71	78±58	79±33

Table 1. Mean and standard deviation of accuracy and agreement measure for the sleep staging obtained by RLS and wavelet feature extraction methods for LD classifier. Acc means general accuracy, kappa is kappa index, Se is the sleep efficiency, REM%TST is the percentage of Total Sleep Time (TST) occupied by REM stages, and Lat represents the latency, in minutes, of the first REM cycle. "Auto" suffixed to one of the previous abbreviations means that the parameter is obtained by the automatic system, despite the suffix "hyp" relates parameters obtained from standard hypnogram.

Quadratic Discriminant

The quadratic discriminant has the best performance on each classifier using RLS based features, and the optimal values were obtained using all the 17 recordings chosen above. It allows for a mean classification kappa index value of 0.55 and accuracy of 0.7680. This method identifies first REM latency with a good precision, indicating a sleep quality close to medical prediction. Table 2 shows comparisons between the two feature extraction

methods. Wavelet method can provide a good estimation of sleep efficiency, but it's worse than RLS classification as regards the other parameters.

	QD classification							
	Acc	**kappa**	**SeAuto**	**SeHyp**	**REM%TST auto**	**REM%TST hyp**	**Lat Auto**	**Lat Hyp**
RLS	76.81 ± 7.51	0.55 ± 0.10	90.25 ± 4.67	85.21 ± 7.25	30.02 ± 14.49	23.42 ± 6.71	86 ± 52	79 ± 33
Wavelet	69.27 ± 12.64	0.35 ± 0.15	91.65 ± 3.25	85.21 ± 7.25	19.36 ± 25.54	23.42 ± 6.71	104±85	79 ± 33

Table 2. Mean and standard deviation of accuracy and agreement measure for the sleep staging obtained by RLS and wavelet feature extraction methods for QD classifier.

K-Nearest-Neighbors

The optimal parameter for this classifier is *K*= 25 applied to features extracted with RLS method, leading to an average kappa index value of 0,42 and an average accuracy of 72%. It can be noticed from Table 3 how this classifier has not a better performance in identifying the sleep stages than other classifiers, especially in identifying clinical parameters. The good classification of sleep efficiency could be considered almost casual, basing on other parameters.

	K-NN classification							
	Acc	**kappa**	**SeAuto**	**SeHyp**	**REM%TST auto**	**REM%TST hyp**	**Lat Auto**	**Lat Hyp**
RLS	71.95 ± 7.47	0.42 ± 0.10	78.77 ± 10.06	85.21 ± 7.25	6.80 ± 8.02	23.42 ± 6.71	201±149	79 ± 33
Wavelet	64.75 ± 9.30	0.33 ± 0.10	64.55 ± 14.63	85.21 ± 7.25	0.75 ± 1.10	23.42 ± 6.71	239±122	79 ± 33

Table 3. Mean and standard deviation of accuracy and agreement measure for the sleep staging obtained by RLS and wavelet feature extraction methods for K-NN classifier.

Feed-Forward Neural Network

Table 4 shows results of Feed-Forward Neural Network (FFNN) based on back-propagation algorithm analysis. One can observe that accuracy and kappa index are worse than the ones obtained with the other classifiers, but also that analysis based on RLS features presents a good sleep efficiency prediction and that wavelet analysis returns good estimation of REM%TST and of first REM cycle latency. However these results are not consistent because of lower total accuracy and lower total kappa index. It is clear that

FFNN not perform good sleep staging by looking single predicted sleep profiles, they often confuse WAKE and REM stages with both signal processing methods.

	FFNN classification							
	Acc	**kappa**	**SeAuto**	**SeHyp**	**REM%TST auto**	**REM%TST hyp**	**Lat Auto**	**Lat Hyp**
RLS	67.17 ± 11.88	0.39 ± 0.13	71.60 ± 17.62	85.21 ± 7.25	11.60 ± 14.14	23.42 ± 6.71	200±141	79 ± 33
Wavelet	50.37 ± 18.79	0.24 ± 0.13	48.75 ± 23.47	85.21 ± 7.25	22.08 ± 28.31	23.42 ± 6.71	87±122	79 ± 33

Table 4. Mean and standard deviation of accuracy and agreement measure for the sleep staging obtained by RLS and wavelet feature extraction methods for FFNN classifier.

Concluding remarks

The performances obtained show that sleep profile can be reached in a very reliable way using bed sensor signals. The LD or QD classifier, combined with a features extraction method based on RLS analysis, can be used to implement a simple automatic system to evaluating sleep macrostructure from time series acquired from bed sensor. These classifiers offer interesting possibility to evaluate sleep efficiency, REM%TST, and first REM cycle latency that are indexes of sleep quality. The obtained results can also demonstrate that a feature extraction method based on RLS model is more preferable than the method based on wavelet decomposition. We expected better results than the obtained ones in *K*-NN and FFNN analysis; this is caused perhaps by an LD-based feature selection. Future researches can explore the possibility of increment *K*-NN and FFNN performances by a dedicated feature selection.

These results are in line with current literature. Redmond *et. al* [17] have presented an automatic algorithm to detect sleep profile from cardiorespiratory signals. The introduction of movement activity allowed an increment of performances from k=0.47 in Redmond *et. al* study to k=0.54 in this study using LD classifier.

INDEX

Chapter 1

Introduction

During the last years, the importance of sleep evaluation has increased due to a considerable number of pathologies that implies sleep disorders. Furthermore, the performance of many basic activities in the normal life such as memorization, learning, productivity and concentration, are closely connected to a good sleep quality [7] [78] [79] [80]. In addition to the sociological and physiological consequences produced by low sleep quality, the sleep evaluation is a time consuming task that has to be done by expert clinicians. This evaluation consists in defining different sleep stages through visual scoring of the polysomnography (PSG). PSG includes the recording of many signals such as electroencephalography (EEG), electromyography (EMG), electrocardiography (ECG), electrooculagram (EOG), pulse oximetry and respiration. With the PSG procedure it is possible to observe sleep efficiency, sleep quality [7] and sleep disorders [80]. Although the PSG is an accurate procedure, some inconveniences rise such as specific equipment, dedicated sleep centers and specialized and trained personnel. All these PSG requirements have generated underestimation of the sleep disorders and low accessibility for the general population. Thus, the development of unattended and portable monitoring systems could be of great help. However, until a few years ago, the sleep evaluation was available only in sleep centers; this situation was mainly generated by technological limitations. Nowadays, these limitations have been partially overcome with the introduction of new technologies that allow the acquisition of physiological signals with high precision in different environments [81] [82] [83][84]. In sleep medicine, the standard practice is to divide the sleep time in epochs with length of 30 seconds, and based on the EEG, EOG and EMG behavior each epoch can be scored as stage 1, 2, 3, 4 and REM (Rapid Eye Movement) and Wake. The representing plot of the tagged epochs is called hypnogram. Generally stages 1 to 4 are merged in one stage named NREM (Non

REM). The intrinsic dynamic of NREM-REM-Wake stages (here macrostructure) carries valuable information about sleep quality since this is highly modified during sleep disorders. Thus, a system able to evaluate this structure over sleep time from signals of simple measurement can contribute to fast and accurate sleep screening. ECG can be recorded at home more efficiently than most of the PSG signals. ECG has been applied for sleep modelling in many studies by calculating firstly the heart rate variability (HRV) out of R-R intervals (ECG RRI), and then analyzing the different spectral components of HRV [85] [44] [88]. HRV also gives information about the cardiovascular and respiratory functions during sleep [85] [44]. These interesting characteristics of HRV can be analysed with automatic procedures [17] and thus the HRV analysis can constitute the basis for sleep evaluation at home. New non-invasive systems and procedures to study the cardiorespiratory behavior during sleep have been published [18] [86]. Ballistocardiographic (BCG) signal can be recorded unobtrusively through a pressure sensitive sensor integrated into the bed mattress, and sleep analysis can be done e.g. with one-channel SCSB method by Alihanka et.al., which is based on the extracted body movements, respiration and heart beat signals [87]. Accuracy of the extraction of the heart beat interval (HBI) from BCG signals can be improved by using multichannel bed sensor and spectral averaging [81], which enables us to make HRV based sleep modeling with a bed sensor, using similar algorithms as in the previous studies with the standard ECG recordings [86]. A non-contact bed sensor is easy to use, as it does not require placement of sensors on the subjects body, and it enables sleep monitoring also at home. In addition to the analysis of the HRV obtained from the HBI, the bed sensor can also measure information related to body movements and respiration [81].

This thesis describes a system to automatically recognize the sleep macrostructure from HBI, respiratory and body movement signals. The sleep stages are automatically detected using a time varying autoregressive model and wavelet decomposition as feature extractor, and several classifiers such as Linear and Quadratic Discriminant, feed-forward Artifical Neural Network and *K*-Nearest-Neighobor. Firstly, this study compares the feature extraction based on autoregressive model and discrete wavelet decomposition and secondly, the automatic sleep classification is compared with the expert visual sleep scoring of PSG recordings for each classifier and for each feature extractor.

This thesis work is divided in six chapters, including the current introduction.

The Second Chapter illustrates the main physiological aspects connected with sleep: the significative importance of HRV during sleep time and the macrostructure of sleep are described in order to introduce the new approach to sleep classification.

The Third Chapter contains a general explanation of the mathematical techniques applied both in the data analysis procedure and in the development of the automatic classification: features extractor based on RLS method and wavelet coefficients, statistical analysis, data processing techniques and the methods used for the classification as Linear and Quadratic Discriminant, *K*-Nearest-Neighbor and feed-forward Neural Network.

The Fourth Chapter contains the first experimental part: the parameters used for features extraction (model order and forgetting factor for RLS model, mother wavelet and level decomposition for wavelet method), the choice of features selection algorithm and data processing.

The Fifth Chapter contains the second experimental part reporting its results and the respective comments obtained applying the methods described above.

The Sixth and final chapter provides a conclusive analysis, underlying the best performance obtained and comparing it with the techniques used.

Chapter 2

Sleep

"O Sleep, rest of all things,
mildest of the gods, balm of the soul..."

Ovid

Sleep is defined as a state of unconsciousness from which a person can be aroused. In this state, the brain is relatively more responsive to internal stimuli than external stimuli. Sleep should be distinguished from coma since coma is an unconscious state from which a person cannot be aroused. Sleep is essential for the normal and healthy functioning of the human body. It is a complicated physiological phenomenon that scientists do not fully understand.

Historically, sleep was thought to be a passive state. However, sleep is now known to be a dynamic process, and our brains are active during sleep. Sleep affects our physical and mental health, and is essential for the normal functioning of all the systems of our body, including the immune system. The effect of sleep on the immune system affects one's ability to fight disease and endure sickness.

States of brain activity during sleep and wakefulness result from different activating and inhibiting forces that are generated within the brain. Neurotransmitters (chemicals involved in nerve signaling) control whether one is asleep or awake by acting on nerve cells (neurons) in different parts of the brain. Neurons located in the brainstem actively produce sleep by inhibiting other parts of the brain that keep a person awake.

Animal studies have shown that sleep is necessary for survival. For example, the normal life of rats spans for 2-3 years. However, rats deprived of sleep live for only about 3 weeks develop abnormally low body temperatures and sores on their tails and paws. The sores are probably produced by impairment of the rats' immune system.

In humans, the metabolic activity of the brain decreases significantly after 24 hours of sustained wakefulness. Sleep deprivation results in a decrease in body temperature, a decrease in immune system function as measured by white blood cell count, and a decrease in the release of growth hormone. Sleep deprivation can also cause increased heart rate variability.
For our nervous systems to work properly, sleep is needed. Sleep deprivation makes a person drowsy and unable to concentrate the next day. It also leads to impairment of memory and physical performance and reduced ability to carry out mathematical calculations. If sleep deprivation continues, hallucinations and mood swings may develop.
Release of growth hormone in children and young adults takes place during deep sleep. Most cells of the body show increased production and reduced breakdown of proteins during deep sleep. Sleep helps humans maintain optimal emotional and social functioning while we are awake by giving rest during sleep to the parts of the brain that control emotions and social relations [1].

2.1 Outline of physiology

Due to the close link between heart rhythm and nervous system, this chapter gives an introduction to physiology of the heart. Then, the following concepts will be analyzed:

1. Nervous System: S.N. Central and S.N. Peripheral
2. Autonomic Nervous System (ANS);
3. Sleep-wake cycle or circadian cycle;
4. The macrostructure of sleep.

2.1.1 Nervous System

The nervous system is defined by the presence of a special type of cell, the neuron. Neurons are distinctive in a number of ways, but their most fundamental property is that they communicate with other cells via synapses, which are membrane-to-membrane junctions containing molecular machinery that allows rapid transmission of signals, either electrical or chemical [3].
The nervous system is divided into the central nervous system (CNS) and the peripheral

nervous system (PNS) (Figure 2.1 [2]).

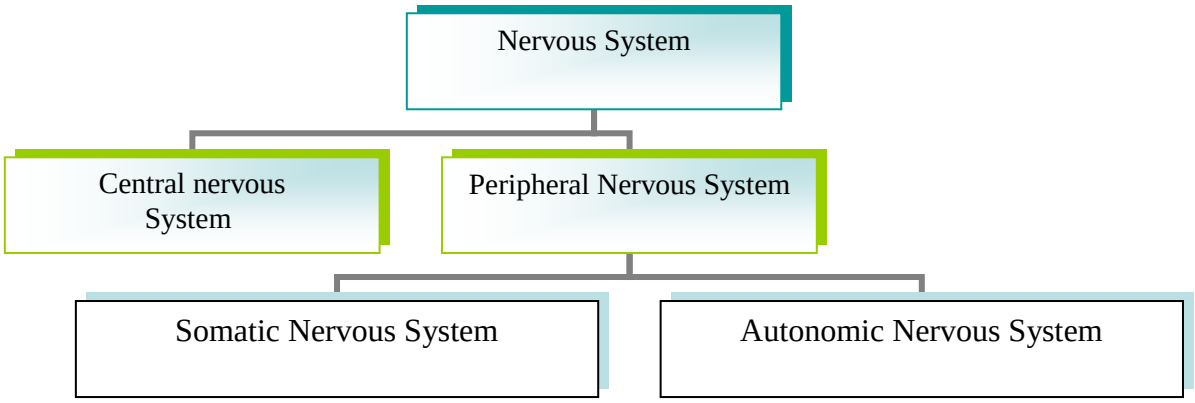


Figure 2.1: Nervous System Division

The central nervous system (CNS) is the part of the nervous system that functions to coordinate the activity of all parts of the body. The central nervous system is enclosed in the meninges. It contains the majority of the nervous system and consists of the brain and the spinal cord.. The CNS is responsible for the integration, analysis and coordination of sensory information and motor commands. CNS is also home to the most important functions such as intelligence, memory, learning and emotions. Unlike the peripheral nervous system, the CNS is not only able to collect and transmit information, but also for integration [1].

There are two types of cells in the peripheral nervous system. These cells carry information to (sensory nervous cells) and from (motor nervous cells) the central nervous system (CNS). Cells of the sensory nervous system send information to the CNS from internal organs or from external stimuli. Motor nervous system cells carry information from the CNS to organs, muscles, and glands. The motor nervous system is divided into the somatic nervous system and the autonomic nervous system. The somatic nervous system controls skeletal muscle as well as external sensory organs such as the skin. This system is said to be voluntary because the responses can be controlled consciously. Reflex reactions of skeletal muscle however are an exception. These are involuntary reactions to external stimuli [4].

On the other hand, the autonomic nervous system controls involuntary muscles, such as smooth and cardiac muscle. This system is also called the involuntary nervous system. The autonomic nervous system can further be divided into the parasympathetic and sympathetic divisions. Figure 2.2 illustrates this division.

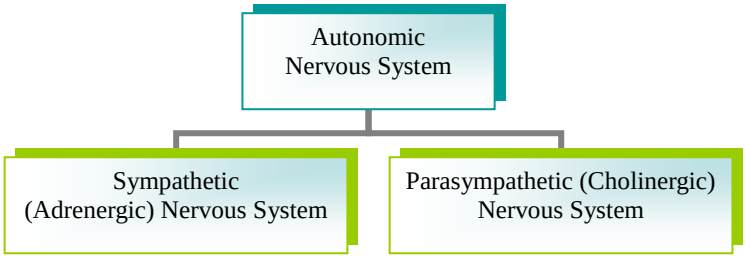


Figure 2.2: Autonomic Nervous System Division

The sympathetic nervous system is frequently referred to as the adrenergic nervous system. Because of its transmitter epinephrine, which is more commonly known by its trade name "Adrenalin", it prepares the body for stress situations. Stimulation of the adrenergic nervous system has the general effect of expending energy. The nerves of the sympathetic division often have an opposite effect when they are located within the same organs as parasympathetic nerves. Nerves of the sympathetic division speed up heart rate, dilate pupils, and relax the bladder. The sympathetic system is also involved in the flight or fight response. This is a response to potential danger that results in accelerated heart rate and an increase in metabolic rate.

The parasympathetic nervous system is usually referred to as the cholinergic nervous system. The cholinergic nervous system is responsible for bringing the body back to normal after the fight or flight response. The effects of the cholinergic nervous system are generally the opposite of those produced by the adrenergic nervous system such as inhibiting heart rate, constricting pupils, and contracting the bladder [4].

The peripheral nervous system is divided into the following sections:

- Sensory Nervous System - sends information to the CNS from internal organs or from external stimuli.
- Motor Nervous System - carries information from the CNS to organs, muscles, and glands.
  - Somatic Nervous System - controls skeletal muscle as well as external sensory organs.
  - Autonomic Nervous System - controls involuntary muscles, such as smooth and cardiac muscle.
    - Sympathetic - controls activities that increase energy expenditures.

- Parasympathetic - controls activities that conserve energy expenditures.

2.1.2 Circadian Rhythm

Biological variations that occur in the course of 24 hours are called circadian rhythms. Circadian rhythms are controlled by the body's biological clock. Many bodily functions follow the biologic clock, but sleep and wakefulness comprise the most important circadian rhythm. Circadian sleep rhythm is one of the several body rhythms modulated by the hypothalamus.

Light directly affects the circadian sleep rhythm. Light is called a *zeitgeber*, a German word meaning time-giver, because it sets the biological clock. A practical purpose has been proposed for the circadian rhythm, using the analogy of the brain being somewhat like a battery charging during sleep and discharging during wakefulness [1].

Body temperature cycles are also under control of the hypothalamus. An increase in body temperature is seen during the course of the day and a decrease is observed during the night. The temperature peaks and troughs are thought to mirror the sleep rhythm. People who are alert late in the evening have body temperature peaks late in the evening, while those who find themselves most alert early in the morning have body temperature peaks early in the evening.

Melatonin, a chemical produced by the pineal gland in the brain, has been implicated as a modulator of light entrainment. It is secreted maximally during the night. Prolactin, testosterone, and growth hormone also demonstrate circadian rhythms, with maximal secretion during the night.

Circadian rhythms can be affected to a certain degree by almost any kind of external stimulus, for example, the beeping of the alarm clock or the timing of meals. When we cross time zones, our circadian rhythms get disrupted leading to jet lag. It usually takes several days for our body rhythms to adjust to the new time.

Symptoms similar to those seen in people with jet lag are common in people who work during nights or work in shifts. Because these people's wake time conflicts with powerful sleep-regulating cues like sunlight, they often become uncontrollably drowsy during work or may have difficulty falling asleep during their off time. Their biological clock wants to do one thing, while they are doing something entirely different. People working in shifts have an increased risk of heart, gastrointestinal, emotional, and mental problems. All these problems

may be related to the disruption of the circadian sleep rhythm [5].

2.1.3 Sleep Macrostructure

As mentioned earlier, sleep is a dynamic process. There are distinct states that alternate in cycles and reflect differing levels of neuronal activity. Each state is characterized by a different type of brain wave activity. Usually the sleep stages are classified in five stages: 1, 2, 3, 4 and REM (rapid eye movement) sleep. These stages progress cyclically from 1 through REM then begin again with stage 1. A complete sleep cycle takes an average of 90 to 110 minutes. The first sleep cycles each night have relatively short REM sleeps and long periods of deep sleep but later in the night, REM periods lengthen and deep sleep time decreases (Figure 2.3).

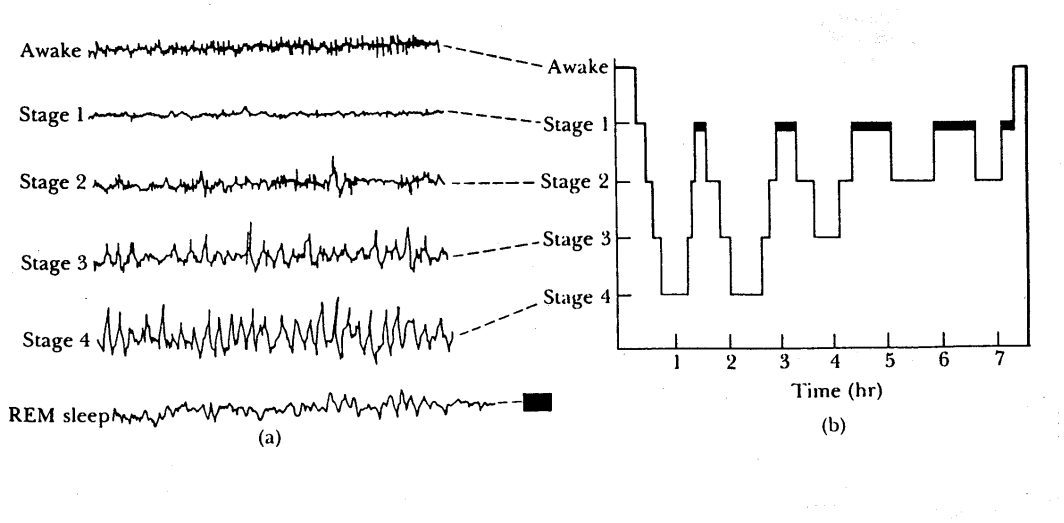


Figure 2.3: Hypnogram with EEG waveforms corresponding to sleep stages.

Sleep stages are divided in five stages:

1. **Stage I** is the beginning of the sleep cycle, and is a relatively light stage of sleep. Stage I can be considered a transition period between wakefulness and sleep. In Stage I, the brain produces high amplitude theta waves, which are very slow brain waves. This period of sleep lasts only a brief time (around 5-10 minutes), and if you awaken someone in the stage, they might report that they weren't really asleep.
2. **Stage II** is the second stage of sleep and lasts for approximately 20 minutes. The brain

begins to produce bursts of rapid, rhythmic brain wave activity known as sleep spindles. Body temperature starts to decrease and heart rate begins to slow [1].

3. **Stage III** is a transitional period between light sleep and a very deep sleep. In this stage, extremely slow brain waves called delta waves (Figure 2.4) begin to appear. They are interspersed with smaller, faster waves.
4. In **stage IV**, delta waves are the primary waves recorded from the brain. These two stages are distinguished from each other only by the percentage of delta activity. Together they represent up to 20% of total sleep time. Stages III and IV are called deep sleep, during which all eye and muscle movement ceases. It is difficult to wake up someone during these two stages. If someone is awakened during deep sleep, he does not adjust immediately and often feels groggy and disoriented for several minutes after waking up. Some children experience bedwetting, night terrors, or sleepwalking during deep sleep [6].
5. Most dreaming occurs during the **fifth stage** of sleep, known as rapid eye movement (REM) sleep. REM sleep is characterized by eye movement, increased respiration rate, increased brain activity. REM sleep is also referred to as paradoxical sleep because while the brain and other body systems become more active muscles become more relaxed. REM sleep represents 20-25% of the total sleep time. REM sleep follows NREM sleep and occurs 4-5 times during a normal 8- to 9-hour sleep period. The first REM period of the night may be less than 10 minutes in duration, while the last may exceed 60 minutes. In a normal night’s sleep, bouts of REM occur every 90 minutes.

When the person is extremely sleepy, the duration of each period of REM sleep is very short or it may even be absent. REM sleep is usually associated with dreaming. Dreaming occurs due because of increased brain activity, but voluntary muscles become paralyzed. The electrical activity recorded in the brain during REM sleep is similar to that which is recorded during wakefulness.

Infants spend almost 50% of their time in REM sleep. Adults spend nearly half of sleep time in stage 2, about 20% in REM and the other 30% is divided between the other three stages. Older adults spend progressively less time in REM sleep (Figure 2.5).

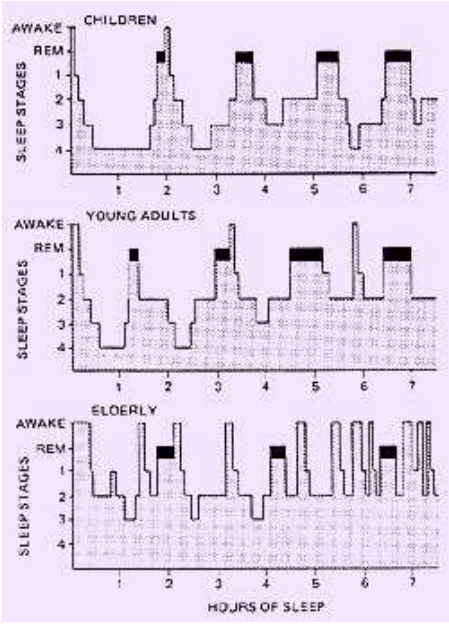


Figure 2.5: Typical hypnograms for children, young adults, and elderly

As sleep research is still a relatively young field, scientists did not discover REM sleep until 1953 when new machines were developed to monitor brain activity. Before this discovery it was believed that most brain activity ceased during sleep. Since then, scientists have also disproved the idea that deprivation of REM sleep can lead to insanity and have found that lack of REM sleep can alleviate clinical depression although they do not know why. Recent theories link REM sleep to learning and memory.

Stage	Frequency (Hz)	Amplitude (microVolts)	Waveform type
Awake	15 – 50	<50	
pre-sleep	8 – 12	50	alpha rhythm
1	4 – 8	50 – 100	theta
2	4 – 15	50 – 150	spindle waves
3	2 – 4	100 – 150	spindle waves and slow waves
4	0.5 – 2	100 – 200	slow waves and delta waves
REM	15-30	<50	

Table 2.1: Summary of waves in the stages

2.2 Polysomnography

Polysomnography is a comprehensive recording of the biophysiological changes that occur during sleep such as (EEG), eye movements (EOG), muscle activity orthoskeletal, muscle activation (EMG) and heart rhythm (ECG) (Figure 2.4).

- The ***electroencephalogram*** (EEG) generally uses six "exploring" electrodes and two "reference" electrodes, unless a seizure disorder is suspected, in which case more electrodes will be applied to document the appearance of seizure activity. The exploring electrodes are usually attached to the scalp near the frontal, central (top) and occipital (back) portions of the brain via a paste that will conduct electrical signals originating from the neurons of the cortex.
- The ***electrooculogram*** (EOG) uses two electrodes; one that is placed 1 cm above the outer canthus of the right eye and one that is placed 1 cm below the outer canthus of the left eye. These electrodes pick up the activity of the eyes in virtue of the electropotential difference between the cornea and the retina (the cornea is positively charged relative to the retina). This determines when REM sleep occurs, of which rapid eye movements are characteristic, and also essentially aids in determining when sleep occurs [8].
- The ***electromyogram*** (EMG) typically uses four electrodes to measure muscle tension in the body as well as to monitor for an excessive amount of leg movements during sleep (which may be indicative of Periodic Limb Movement Disorder, PLMD). Two leads are placed on the chin with one above the jaw line and one below. This, like the EOG, helps determine when sleep occurs as well as REM sleep. Sleep generally includes relaxation and so a marked decrease in muscle tension occurs. A further decrease in skeletal muscle tension occurs in REM sleep. A person becomes partially paralyzed to make acting out of dreams impossible, although people that do not have this paralysis can suffer from REM Behavior Disorder. Finally, two more leads are placed on the anterior tibialis of each leg to measure leg movements.

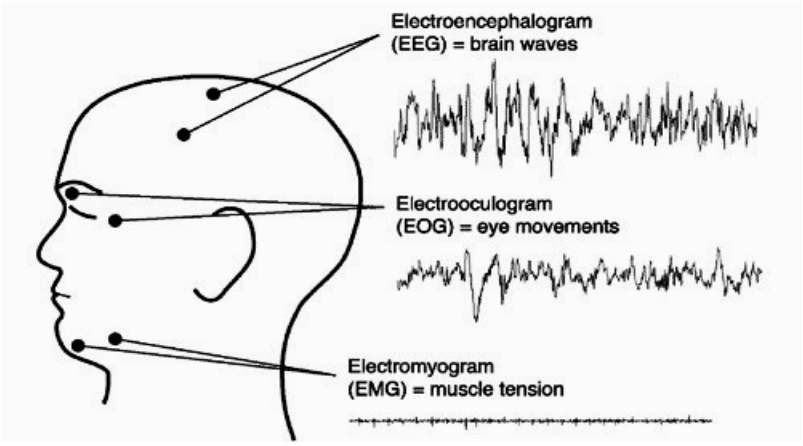


Figure 2.6: Recording of the biophysiological changes that occur during sleep such as (EEG), eye movements (EOG), muscle activity orskeletal, muscle activation (EMG) and heart rhythm (ECG)

- Though a typical ***electrocardiogram*** (ECG or EKG) would use ten electrodes, only two or three are used for a polysomnogram. They can either be placed under the collar bone on each side of the chest, or one under the collar bone and the other six inches above the waist on either side of the body. These electrodes measure the electrical activity of the heart as it contracts and expands, recording such features as the "P" wave, "QRS" complex, and "T" wave. These can be analyzed for any abnormalities that might be indicative of an underlying heart pathology.

Thanks to these contemporaneous records and to other sensors, which provide related to the chest, abdomen and fingers (for respiration and oxygen saturation) (Figure 2.7), doctors give their clinic assessment of the subject's hypnogram according to the classical method of staging sleep [7]. It is observed easily as a person subjected to this type of analysis for an entire night has to stand in conditions far from optimal. Therefore, efforts to evaluate in a simple way the sleep quality could be of benefit for the general population.

Figure 2.7: Recordings which provide to sleep staging

2.3 The electrocardiogram

The electrocardiogram (ECG) is a diagnostic tool that measures and records the electrical activity of the heart. Interpretation of ECG allows diagnosis of a wide range of heart conditions. These conditions can vary from minor to life threatening.

The term electrocardiogram was introduced by Willem Einthoven in 1893 at a meeting of the Dutch Medical Society. In 1924, Einthoven received the Nobel Prize for his life's work in developing the ECG.

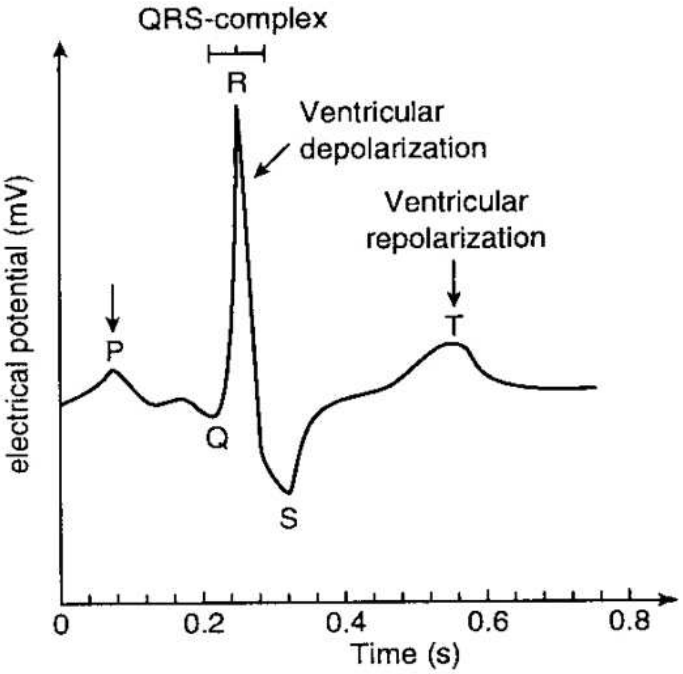


Figure 2.8: Electrocardiogram

The heart normally beats between 60 and 100 times per minute, with many normal variations. For example, athletes at rest have slower heart rates than most people. This rate is set by a small collection of specialized heart cells called the sinoatrial (SA) or sinus node. Located in the right atrium, the sinus node is the heart's "natural pacemaker". The ECG records the electrical activity that results when the heart muscle cells in the atria and ventricles contract (Figure 2.8). The regular pattern of the ECG can be divided into five parts:

- the P wave is caused by the contraction of the atria (atrial systole), which provides indications of the time taken by the impulse to propagate to both atria (can be used precisely for the diagnosis of atrial disease);
- the PQ isoelectric waveless segment indicates the passage of the impulse from the atria to the ventricles;
- QRS complex is formed by the short and negative wave Q, the high and narrow R wave and small S wave, which was also negative. The complex indicates ventricular systole with the arrival of the impulse to the ventricles (Q wave) and their extension to the whole tissue (R and S);
- the long ST interval following the QRS complex and including the T wave can detect ischemic problems. It represents the period during which the ventricles contract and then (with the T wave) back to sleep.

2.4 Heart Rate Variability

Heart rate variability (HRV) describes the variations between consecutive heartbeats. The rhythm of the heart is controlled by the sinoatrial (SA) node, which is modulated by both the sympathetic and parasympathetic branches of the autonomic nervous system. Sympathetic activity tends to increase heart rate (HR↑) and its response is slow (few seconds). Parasympathetic activity, on the other hand, tends to decrease heart rate (HR↓) and mediates faster (0.2–0.6 seconds). In addition to central control, there are some feedback mechanisms that can provide quick reflexes. This reflex is based on baroreceptors which are located on the walls of some large vessels and can sense the stretching of vessel walls caused by pressure increase. Both sympathetic and parasympathetic activity are influenced by baroreceptor stimulation trough a specific baroreflex arc (Figure 2.9). The changes recorded in the HRV signal are determined by mechanical, neuronal, humoral and thermal modifications [9].

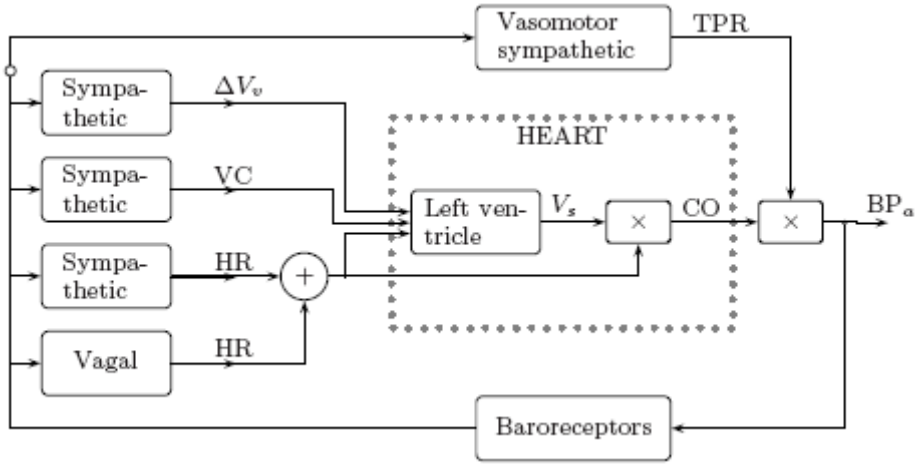


Figure 2.9. The four baroreflex pathways (redrawn from [49]). Variation in venous volume (_Vv), left ventricular contractility (VC), sympathetic and parasympathetic (vagal) control of heart rate (HR), stroke volume (Vs), cardiac output (CO), total peripheral resistance (TPR), and arterial blood pressure (BPa).

The mathematical transformation (Fast Fourier Transform) of HRV data into power spectral density (PSD) is used to discriminate and quantify sympathetic and parasympathetic activity and total autonomic nervous system activity. Power spectral analysis reduces the HRV signal into its constituent frequency components and quantifies the relative power of these components [11]. The continuous modulation of the sympathetic and parasympathetic innervations results in variations in heart rate.

The power spectrum is divided into three main frequency ranges:

- the very low frequency range (VLF) (0.0033 to 0.04 Hz), representing slower changes in heart rate, is an index of sympathetic activity;
- the low frequency range (LF) (0.04 to 0.15) is also often referred to as the baroreceptor band, because it reflects the blood pressure feedback signals sent from the heart back to the brain, which also affect the HRV waveform. The LF band is more complex, as it can reflect a mixture of sympathetic and parasympathetic activity;
- the high frequency range (HF) (0.15 to 0.4 Hz), representing quicker changes in heart rate, is primarily due to parasympathetic activity;

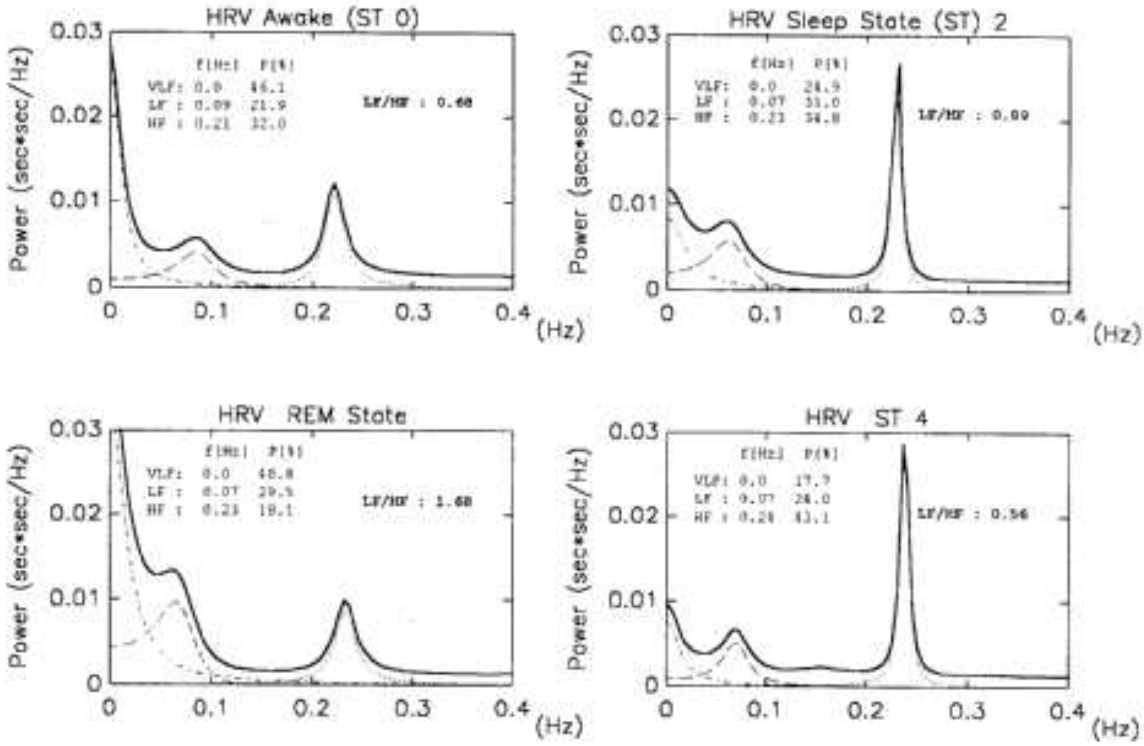


Figure 2.10: Power spectral density of HRV signal during some phases of sleep.

In Stages III and IV (slow-wave sleep), it is observed an increase in the HF peak, so sleep is synchronized, so ratio LF / HF more decreases. During REM, as well as having an increase of LF / HF, there is a more wide distribution of peak HF, due to irregular breathing. In HRV spectral analysis through the entire night, in the latter stages REM, sympathetic activation reaches values higher than the state of wakefulness and frequencies which are centered around the LF and HF peaks are lowered [10] [12]. These considerations led to the idea of being able to use the HRV signal for classify the different stages of sleep (Figure 2.10).

Even though HRV has been studied extensively during the last decades within which numerous research articles have been published, the practical use of HRV have reached

general consensus only in two clinical applications [44]. That is, it can be used as a predictor of risk after myocardial infarction [47, 48] and as an early warning sign of diabetic neuropathy [45,46]. In addition, HRV has been found to correlate with, e.g., age, mental and physical stress, and attention, see e.g., a review in [42].

The starting point for HRV analysis is the ECG recording from which the HRV time series can be extracted. In the formulation of the HRV time series, a fundamental issue is the determination of heartbeat rate.

2.4.1 Heart beat period and QRS detection

The aim in HRV analysis is to examine the sinus rhythm modulated by the autonomic nervous system. Therefore, one should technically detect the occurrence times of the SA-node action potentials. This is, however, practically impossible and, thus, the fiducial points for the heart beat is usually determined from the ECG recording. The nearest observable activity in the ECG compared to SA-node firing is the P-wave resulting from atrial depolarization (Figure 2.11) and, thus, the heart beat period is generally defined as the time difference between two successive P-waves. The signal-to-noise ratio of the P-wave is, however, clearly lower than that of the strong QRS complex which results primarily from ventricular depolarization. Therefore, the heart beat period is commonly evaluated as the time difference between the easily detectable QRS complexes.

A typical QRS detector consists of a pre-processing part followed by a decision rule. Several different QRS detectors have been proposed within last decades [50, 51, 52, 53, 54]. For an easy to read review of these methods, see [55]. The pre-processing of the ECG usually includes at least bandpass filtering to reduce power line noise, baseline wander, muscle noise, and other interference components. The bandpass can be set to approximately 5–30 Hz which covers most of the frequency content of QRS complex [51]. In addition, pre-processing can include differentiation and/or squaring of the samples. After pre-processing, the decision rules are applied to determine whether or not a QRS complex has occurred. The decision rule usually includes an amplitude threshold which is adjusted adaptively as the detection progresses. In addition, the average heart beat period is often used in the decision. The fiducial point is generally selected to be the R-wave and the corresponding time instants are given as the output of the detector.

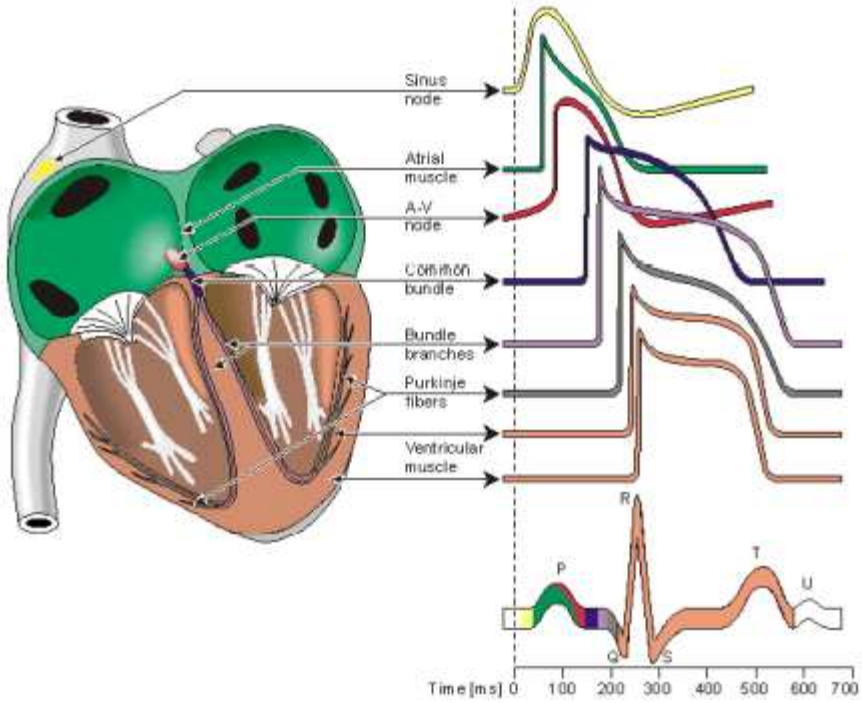


Figure 2.11: Electrophysiology of the heart (redrawn from [68]). The different waveforms for each of the specialized cells found in the heart are shown. The latency shown approximates that normally found in the healthy heart.

The accuracy of the R-wave occurrence time estimates is often required to be 1–2 ms and, thus, the sampling frequency of the ECG should be at least 500–1000 Hz [44]. If the sampling frequency of the ECG is less than 500 Hz, the errors in R-wave occurrence times can cause critical distortion to HRV analysis results, especially to spectrum estimates [56]. The distortion of the spectrum is even bigger if the overall variability in heart rate is small [57]. The estimation accuracy can however be improved by interpolating the QRS complex (for example by using a cubic spline interpolation [58] or some model based approach [59]). It should be, however, noted that when the SA-node impulses are of interest there is an unavoidable estimation error of approximately 3 ms due to fluctuations in the AV-nodal conduction time [60].

2.4.2 Derivation of HRV time series

After the QRS complex occurrence times have been estimated, the HRV time series can be derived. The inter-beat intervals or RR intervals are obtained as differences between successive R-wave occurrence times. That is, the n^{th} RR interval is obtained as the difference

between the R-wave occurrence times $RR_n = t_n - t_{n-1}$. In some context, normal-to-normal (NN) may also be used when referring to these intervals indicating strictly intervals between successive QRS complexes resulting from SA-node depolarization [44]. In practice, the NN and RR intervals appear to be the same and, thus, the term RR is preferred here. The time series constructed from all available RR intervals is, clearly, not equidistantly sampled, but has to be presented as a function of time, i.e. as values (t_n, RR_n). This fact has to be taken into account before frequency-domain analysis. In general, three different approaches have been used to get around this issue [44]. The simplest approach that has been adopted in, e.g., [61] is to assume equidistant sampling and calculate the spectrum directly from the RR interval tachogram (RR intervals as a function of beat number), see the first panel of Figure 2.12.

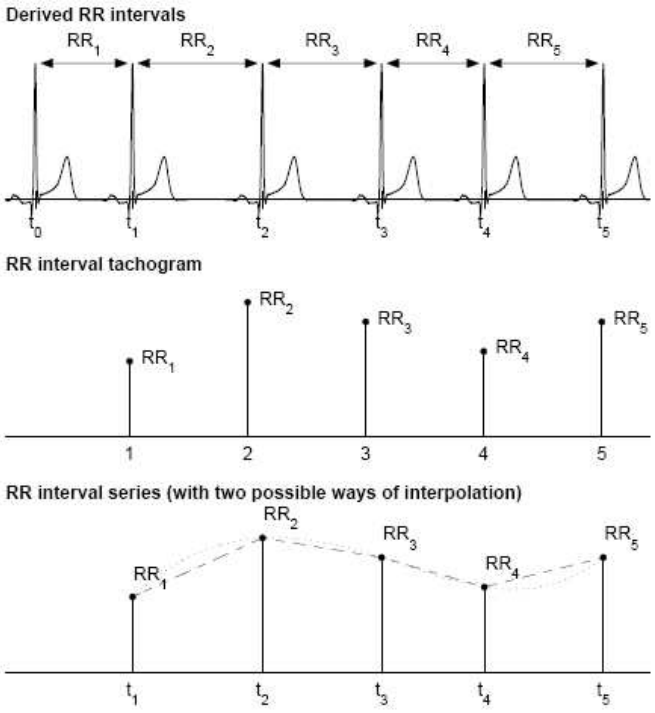


Figure 2.12: Derivation of two HRV signals from ECG: the interval tachogram (middle panel) and interpolated RR interval series (bottom panel).

This assumption can, however, cause distortion into the spectrum [62]. This distortion becomes substantial when the variability is large in comparison with the mean level. Furthermore, the spectrum can not be considered to be a function of frequency but rather of cycles per beat [63]. Another common approach, adopted in this software, is to use interpolation methods for converting the non-equidistantly sampled RR interval time series (also called the interval function) to equidistantly sampled [44], see the bottom panel of Figure 2.12. One choice for the interpolation method is the cubic spline or linear interpolation

[62]. After interpolation, regular spectrum estimation methods can be applied. The third general approach called the spectrum of counts considers a series of impulses (delta functions positioned at beat occurrence times) [64]. This approach relies on the generally accepted integral pulse frequency modulator (IPFM) which aims to model the neural modulation of the SA-node [60]. According to this model, the modulating signal is integrated until a reference level is achieved after which an impulse is emitted and the integrator is set to zero. The spectrum of the series of events can be calculated, e.g., by first lowpass filtering the event series and then calculating the spectrum of the resulting signal [63].

2.4.3 Pre-processing of HRV time series

Any artefact in the RR interval time series may generate problems in the analysis of these signals. The artefacts within HRV signals can be divided into technical and physiological artefacts. The technical artefacts can include missing or additional QRS complex detections and errors in R-wave occurrence times. These artefacts may be due to measurement system or the computational algorithm. The physiological artefacts, on the other hand, include ectopic beats and arrhythmic events. In order to avoid the interference of such artefacts, the ECG recording and the corresponding event series should always be manually checked for artefacts and only artefact-free sections should be included in the analysis [44]. Alternatively, if the amount of artefact-free data is insufficient, proper interpolation methods can be used to reduce these artefacts, see, e.g., [65, 66, 67].

Another common feature that can alter the analysis significantly is the slow linear or more complex trends within the analyzed time series. Such slow non-stationarities are characteristic for HRV signals and should be considered before the analysis. The origins of non-stationarities in HRV are discussed, e.g., in [42].

2.5 Automatic classification of sleep stages

The computer support can be very useful in the case of sleep medicine and particularly in its automatic assessment. The present research follows this direction. Sleep classification needs a massive amount of data due to registration of different signals along an entire night. It needs to be made in sleep centers with a specialized team of doctors and a group of technical

experts. Major obstacle to diagnosis in the field of sleep is represented by sleep centers because they are a very few and procedures for sleep evaluation are very expensive. Thus, there are large waiting lists and this could generate some negligence in estimating of major sleep disorders. It was thought to introduce into the sleep study mathematical models and computer resources, thanks to their continued development and their ability to manage these data. In particular, Decision Support Systems (DSS) are useful to reduce substantially the work of experts. This research moves into two main directions: on the one hand, sleep stages are classified by the signals usually used by doctors in scoring (especially using the EEG), on the other hand we will try to automatically mimic the medical score using only the HRV.

In the first view, there is the work of Prince et al. (1989) [13]. He develops a model for the automatic classification of the sleep stages combining the information given by different signals, in other words human knowledge and mathematical structure. Typical signals (EEG / EOG / EMG) were used, which once processed (discredited) were the inputs of an automatic classifier. With his work, Prince has tried to classify automatically the five stages of sleep, but he didn't achieve optimal performance. Other important works are those of Agarwal and Gotman in 2001 (using clusters) [16], Pore in 2006 (Independent Component Analysis) [14] and Flexer in 2005 (study by HMM using only EEG signal) [15].

The second area of research is about Heart Rate Variability. As mentioned, this signal is close physiological sleep and it is also very robust against noise and simple purchase. These features make it particularly suitable for the development of DSS and easily acquiring in house, thus eliminating the problems of cost and time for display (Figure 2.13). Important works in this field are those of Redmond in 2007 which tried to classify the three major phases of sleep (WAKE, REM, NREM) through the use of cardiac and respiratory signals. The stages were calculated using a linear classifier [17].

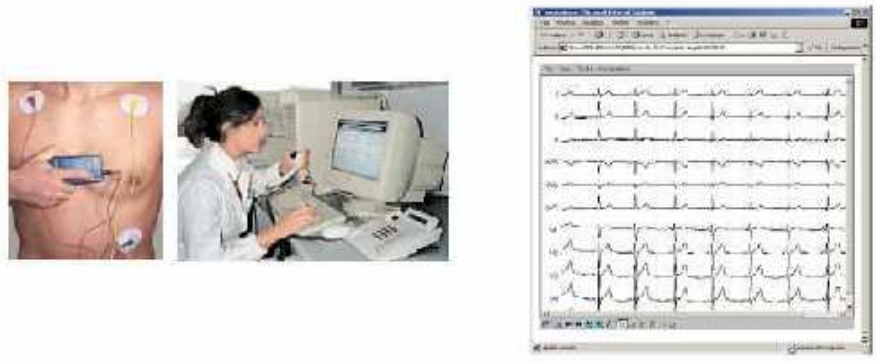


Figure 2.13: ECG signal detection and its transfer to PC.

Other attempts have been made previously by Watanabe in 2004 [18] and more recently by Lewicki (only for two sleep stages) [20]. In the same field of research there is also work of Politecnico di Milano by using HMM (REM-NREM classification) [19].

Chapter 3

Mathematical Methods

"Whether you can observe a thing or not depends on the theory which you use. It is the theory which decides what can be observed."

A. Einstein

This chapter presents the mathematical approaches used to develop this work. The spectral heart rate variability estimation and the statistical analysis during signal processing are presented; afterwards it will follow a general description of some classification and validation techniques used.

3.1 Spectral estimation of the HRV

It was explained as a heart rate variability signal has non-random fluctuations. These fluctuations, around its mean value, represent the sympathetic and vagal balance and are deeply linked to the physiology of sleep. The common used tool is the frequency domain analysis of series extracted from the HRV signal (VLF, LF, HF), which will form the basis for the classification of sleep [23]. The classic analysis is usually carried out by Fourier transform (FFT). It also require a wide temporal window (5 minutes) in order to catch the low frequency bands (such as VLF) while the temporal band considered in this work corresponds to 30 seconds. Fortunately, the Wiener–Khinchin theorem provides a simple alternative. The power spectral density (PSD) is the Fourier transform of the autocorrelation function, $R(\tau)$, of the signal if the signal can be treated as a wide-sense stationary random process [21].

Depending on what is known about the signal, estimation techniques can involve parametric or non-parametric approaches, and may be based on time-domain or frequency-domain analysis. In parametric approaches, the task is to estimate the parameters of the model that describes the stochastic process. For example, a common parametric technique involves fitting the observations by an *autoregressive model*. By contrast, non-parametric approaches explicitly estimate the covariance or the spectrum of the process without assuming that the process has any particular structure. A common non-parametric technique is the *periodogram* [22]. Spectral density is a function of frequency, not a function of time. Sometimes it is useful estimate PSD across the time, then techniques such as the *short-time Fourier transform* and *wavelets*, are commonly used. The present study analyze the application of parametric (RLS) and non parametric approach (Wavelet) as feature extractors for sleep analysis based on HRV signal.

3.1.1 Spectral estimation of non-stationary signals

Analysis of HRV data has traditionally been performed in the time domain. Unfortunately, time analysis often fails to detect the presence of certain rhythms and subtle changes which may be crucial in many applications. This fact has motivated the use of frequency domain techniques for analysis. However, as the HRV can be described as a non-stationary random signal with additive stationary random noise and since only one realization of data is available, accurate time-varying spectral estimates can be extremely difficult to obtain. There are many approaches to estimating the time-varying spectra of a non-stationary random signal. For a random signal the expected value of a time-frequency distribution (TFD) is known as the time-frequency spectra (TFS) [25] or evolutive spectra [26]. If multiple realizations of the signal are available, one can form an estimate of the time-varying spectrum by averaging over a set of TFDs. Where each TFD in the set is formed from a separate realization. When a random signal is neither stationary or at the very least quasi-stationary, ergodicity cannot be assumed, and thus the ensemble average cannot be replaced by the time average. Any time averaging, even over short periods, will smooth the time-varying characteristics of the signal. Thus, when only one realization of the signal is available, as is the case for HRV data, the TFD of a single realization must be used,

followed by a smoothing chosen to be a compromise between the two requirements of obtaining an accurate spectral estimate and preserving the time-varying information. In addition to the single realization problem, most time-varying spectral estimation techniques have no direct means for discriminating between the wanted signal and unwanted noise. Thus the resulting spectral estimate will contain the characteristics of the noise. Parametric methods are an exception to this since they can separate signal and noise structures, allowing for better spectral estimates [27].

In this discuss two known methods for estimating the time-varying spectra of a non-stationary signal are used. One of the methods is based on modelling the signal as a time-varying autoregressive process and using a recursive filter (RLS). The other approach uses the Wavelet Transform which provides a unified framework for a number of methods, which have been developed independently for various signal processing applications [24].

3.1.2 Autoregressive models

The parametric model input, which in general is a white noise, is related output from the following equation [23]:

$$x_n = \sum_{i=0}^{q} b_i y_{n-1} - \sum_{k=1}^{p} a_k x_{n-k} \quad (3.1)$$

The transfer function is given by $H(z)=\frac{B(z)}{A(z)}$, where B(z) represents the z-transform of moving average filter (MA) and A(z) represents autoregressive part (AR). It is an ARMA model or a linear model with terms of moving average and autoregressive terms. There are particular models deriving by the ARMA model, considering only the autoregressive (AR models) or the moving average (model MA). AR model is widely used because it represents an alternative to the Fourier transform and has computationally advantages.

The basic idea behind autoregressive models is that, for certain type of processes, the current value of the variable can be expressed in terms of a combination of a finite number of terms of the past. There is "memory" or feedback and therefore the system can generate

internal dynamics.

$$x_n = \sum_{k=1}^{p} a_k x_{n-k} + \varepsilon_n \quad (3.2)$$

where a_k are autoregression coefficients, x is the series under investigation, and N is the order (length) of the filter which is generally very much less than the length of the series. The noise term or residue, epsilon in the above, is almost always assumed to be Gaussian white noise. Verbally, the current term of the series can be estimated by a linear weighted sum of previous terms in the series. The weights are the autoregression coefficients. The problem in AR analysis is to derive the "best" values for a_k given a series x_n . The majority of methods assume the series is linear and stationary. By convention the series is assumed to be zero mean, if not this is simply another term a_0 in front of the summation in the equation above (3.1). Obtaining the AR coefficients in the stationary-case, usually it minimizes:

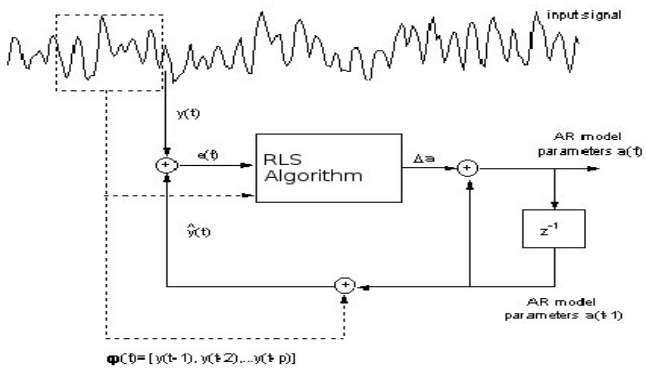
$$J = \frac{1}{N} \sum_{t=1}^{N} [e(t)^2] \quad (3.3)$$

where e(t) represents the prediction error $e(t) = y(t) - \hat{y}$.

When there is a new sample you must obtain a new model adapted to the new data and therefore all calculations must be re-initialising. One way to approach is to use a recursive algorithm: it calculates the prediction error for each new sample, based on this, update the vector of model parameters $\vartheta(t)$ in accordance with the following expression:

$$\vartheta(t) = \vartheta(t-1) + G(t)[y(t) - \hat{y}(t)] \quad (3.4)$$

The terms of prediction error have equal weight in the case of stationary signals. In the case of transient signals, however, is appropriate to update them to match the new data gradually observed, giving more weight samples more recent and less weight to the previous ones (Figure 3.1) [29].



$$\mathbf{a}(t) = \mathbf{a}(t-1) + \mathbf{K}(t)e(t)$$

$$e(t) = y(t) - \hat{y}(t)$$

$$\hat{y}(t) = \varphi^{T}(t)\mathbf{a}(t-1)$$

$$\mathbf{K}(t) = \frac{\mathbf{P}(t-1)\varphi(t)}{\lambda + \varphi(t)^{T}\mathbf{P}(t-1)\varphi(t)}$$

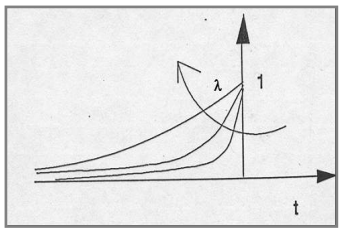
$$\mathbf{P}(t) = \frac{1}{\lambda}\left[\mathbf{P}(t-1) - \frac{\mathbf{P}(t-1)\varphi(t)\varphi(t)^{T}\mathbf{P}(t-1)}{\lambda + \varphi(t)^{T}\mathbf{P}(t-1)\varphi(t)}\right]$$

Figure 3.1: RLS model used for the analysis of data from the signal HRV. At the top you look at the temporal window used for the analysis of samples.

Now it minimizes:

$$J = \frac{1}{N}\sum_{t=1}^{N} \omega^{N-t}[e(t)^{2}] \quad (3.5)$$

The term ω represents the *forgetting factor* (Figure 3.2) that allows to obtain a sequence of error weighting window, upgrading significantly to new data and less to the old. When the value of this factor equals 1, the window is rectangular and then it is equivalent to stationarity-case; when ω<1 the window is more small and therefore the number of error terms that contribute to *H* model becomes: $\frac{1}{1-\omega}$. The choice of ω is crucial for the estimation of the model, in fact while a too low ω could estimate significantly noise; ω high values make the algorithm unstable and unable to catch signal variability. In this work as parametric model it uses a recursive least square (RLS Figure 3.1 [28]). Following the HRV signal, the better choice of this parameter is 0.985, after some trials.



$$N = \frac{1}{1-\lambda} \quad \begin{cases} \lambda = 1 & \text{All the samples} \\ \lambda = 0.99 & \text{100 samples} \\ \lambda = 0.98 & \text{50 samples} \end{cases}$$

Figure 3.2: Forgetting factor and its windowing

The RLS model uses 8 parameters following HRV data; so there are a parametric vector θ and an observation vector φ with the same dimension, as defined:

$$\varphi(t) = [s(t-1)s(t-2)...s(t-8)] \quad (3.6)$$

where

$$s(t) = 0.95s(t-1) + 0.975x(t) - 0.975x(t-1) \quad (3.7)$$

and *x(t)* is model input, HRV signal, filtering through high-pass filter obtaining *s(t)*. Input signal is filtered step-by-step and parameter vector is updated at each step calculating prediction error:

$$\varepsilon(t) = s(t) - \varphi(t)\vartheta(t-1) \quad (3.8)$$

The parameter vector updates as:

$$\vartheta(t) = \vartheta(t-1) + K\varepsilon(t) \quad (3.9)$$

The updating factor K is derived by regularized covariance matrix:

$$P(t) = \frac{P(t-1) + \dfrac{P(t-1)\varphi(t)^T \varphi(t)P(t-1)}{\omega + \varphi(t)P(t-1)\varphi(t)^T}}{\omega} \quad (3.10)$$

through the relationship:

$$K(t) = P(t)\varphi(t)^T \quad (3.11)$$

The RLS initialising is:

$$\begin{aligned} P(0) &= I \\ \vartheta(0) &= [1 \quad 1 \quad ... \quad 1] \end{aligned} \quad (3.12)$$

So each time point has optimal parametric vector, minimizing prediction error variance. The signal variance is estimated recursively:

$$\sigma^2(t) = \varepsilon^2(t)(1-\omega) + \sigma^2(t-1)\omega \quad (3.13)$$

Spectral estimation is computed in each time as autoregressive model by parametric vector $\vartheta(t)$ and $\sigma^2(t)$:

$$W(f,t) = \frac{\sigma^2(t)}{f_{sample}\left|\sum_{n=1}^{8} \vartheta_n(t) e^{-\pi j n f} + 1\right|^2} \quad (3.14)$$

where:

- f_{sample} is sampling frequency of HRV signal, *x(t)*
- $\theta_n(t)$ is nth parameter of vector $\theta(t)$
- f is normalized frequency between 0 and 1.

In this way, it was calculated the power in different frequency bands (VLF, LF, HF) and total power (TP). These bands are normalized to the total power, avoiding *inter-* and *intra-*subjective variability.

3.1.3 Wavelet Analysis

The fundamental idea behind wavelets is to analyse according to scale. The wavelet analysis procedure is to adopt a wavelet prototype function called an *analysing* wavelet or *mother* wavelet. Any signal can then be represented by translated and scaled versions of the

mother wavelet. Wavelet analysis is capable of revealing aspects of data that other signal analysis techniques such as Fourier analysis miss aspects like trends, breakdown points, discontinuities in higher derivatives, and self- similarity. Furthermore, because it affords a different view of data than those presented by traditional techniques, it can compress or de-noise a signal without appreciable degradation [30].

3.1.3.1 Wavelet vs. Fourier Analysis

In the well-known Fourier analysis, a signal is broken down into constituent sinusoids of different frequencies. These sines and cosines (essentially complex exponentials) are the basis functions and the elements of Fourier synthesis. Taking the Fourier transform of a signal can be viewed as a rotation in the function space of the signal from the time domain to the frequency domain. Similarly, the wavelet transform can be viewed as transforming the signal from the time domain to the wavelet domain. This new domain contains more complicated basis functions called wavelets, mother wavelets or analysing wavelets.
Mathematically, the process of Fourier analysis is represented by the Fourier transform:

$$F(\omega) = \int_{-\infty}^{\infty} f(t) e^{-j\omega t} dt \quad (3.15)$$

which is the sum over all time of the signal *f(t)* multiplied by a complex exponential. The results of the transform are the Fourier coefficients *F(*ω*)*, which when multiplied by a sinusoid of frequency ω, yield the constituent sinusoidal components of the original signal.
A wavelet prototype function at a scale *s* and a spatial displacement *u* is defined as:

$$\psi_{s,u} = \sqrt{s}\psi\left[\frac{x-u}{s}\right] \quad (3.16)$$

Replacing the complex exponential in (3.12) with this function yields the *continuous wavelet transform* (CWT):

$$C(s,u) = \int_{-\infty}^{\infty} f(t)\sqrt{s}\psi\left[\frac{x-u}{s}\right] \quad (3.17)$$

which is the sum over all time of the signal multiplied by scaled and shifted versions of the wavelet function ψ. The results of the CWT are many wavelet coefficients C, which are a function of scale and position. Multiplying each coefficient by the appropriately scaled and shifted wavelet yields the constituent wavelets of the original signal. The basis functions in both Fourier and wavelet analysis are localised in frequency making mathematical tools such as power spectra (power in a frequency interval) useful at picking out frequencies and calculating power distributions. The most important difference between these two kinds of transforms is that individual wavelet functions are localised in space. In contrast Fourier sine and cosine functions are non-local and are active for all time t. This localisation feature, along with wavelets localisation of frequency, makes many functions and operators using wavelets sparse when transformed into the wavelet domain. This sparseness, in turn results in a number of useful applications such as data compression, detecting features in images and de-noising signals.

3.1.3.2 Time-Frequency Resolution

A major draw back of Fourier analysis is that in transforming to the frequency domain, the time domain information is lost. When looking at the Fourier transform of a signal, it is impossible to tell when a particular event took place. In an effort to correct this deficiency, Dennis Gabor (1946) adapted the Fourier transform to analyse only a small section of the signal at a time a technique called windowing the signal [31]. Gabor's adaptation, called the Windowed Fourier Transform (WFT) gives information about signals simultaneously in the time domain and in the frequency domain. To illustrate the time-frequency resolution differences between the Fourier transform and the wavelet transform consider the following figures.

The figure 3.3 shows a windowed Fourier transform, where the window is simply a square wave. The square wave window truncates the sine or cosine function to fit a window of a particular width. Because a single window is used for all frequencies in the WFT, the

resolution of the analysis is the same at all locations in the time frequency plane.

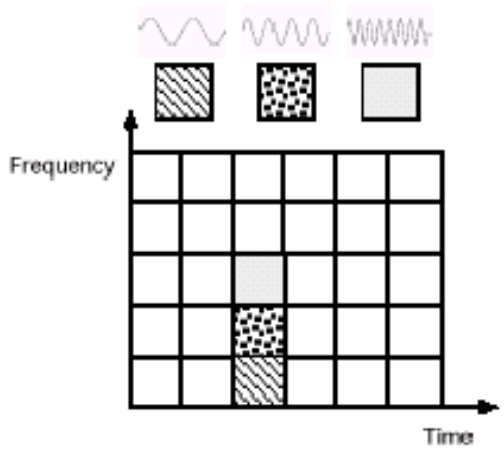


Figure 3.3: Resolution Windowed Fourier Transform

An advantage of wavelet transforms is that the windows vary. Wavelet analysis allows the use of long time intervals where we want more precise low-frequency information, and shorter regions where we want high-frequency information. A way to achieve this is to have short high-frequency basis functions and long low-frequency ones.

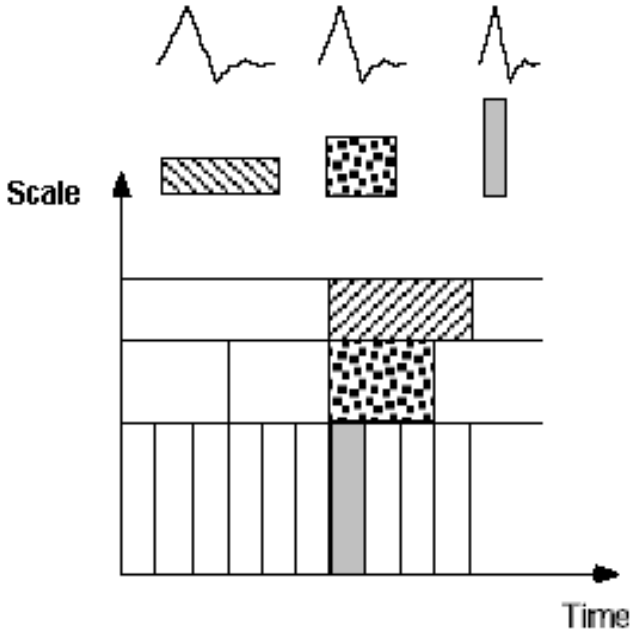


Figure 3.4:Wavelets Resolution

The figure 3.4 shows a time-scale view for wavelet analysis rather than a time frequency

region. Scale is inversely related to frequency. A low-scale compressed wavelet with rapidly changing details corresponds to a high frequency. A high-scale stretched wavelet that is slowly changing has a low frequency.

3.1.3.3 Examples of Wavelets

The figure below illustrates four different types of wavelet basis functions.

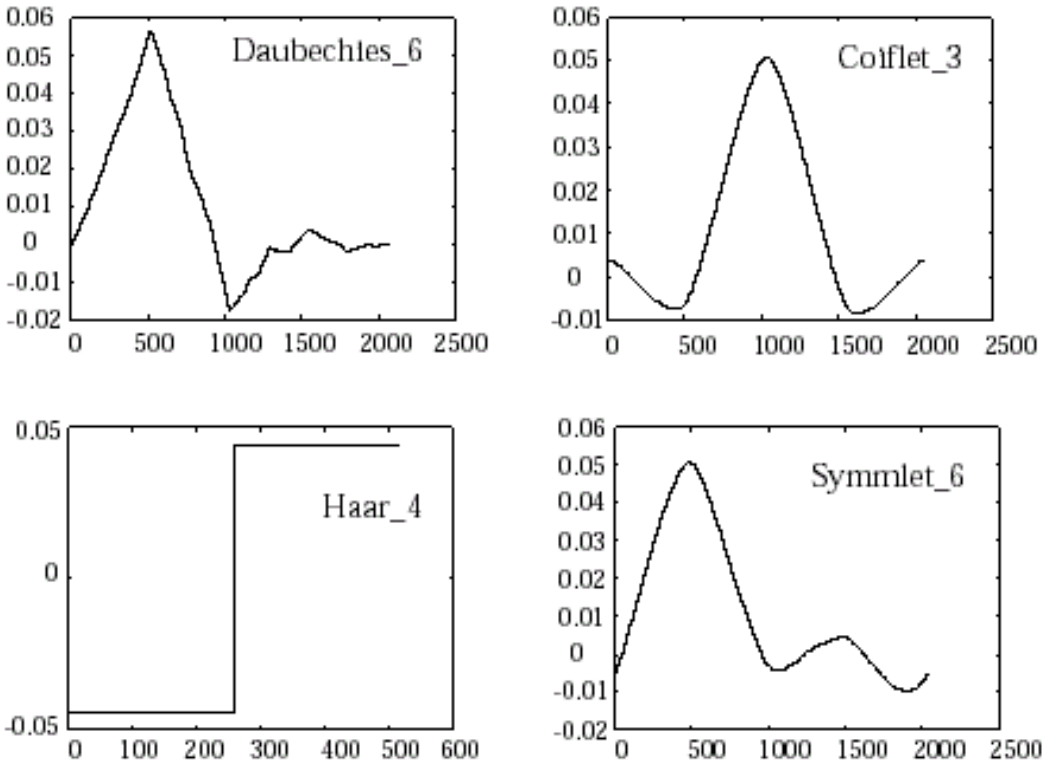


Figure 3.5: Different wavelet families

The different families make trade-offs between how compactly the basis functions are localised in space and how smooth they are. Within each family of wavelets (such as the Daubechies family) are wavelet subclasses distinguished by the number of filter coefficients and the level of iteration. Wavelets are most often classified within a family by the number of *vanishing moments*. This is an extra set of mathematical relationships for the coefficients that must be satisfied. The extent of compactness of signals depends on the number of vanishing moments of the wavelet function used.

3.1.3.4 The Discrete Wavelet Transform

The Discrete Wavelet Transform (DWT) involves choosing scales and positions based on powers of two so-called dyadic scales and positions. The mother wavelet is rescaled or dilated by powers of two and translated by integers. Specifically, a function $f(t) \in L^2(R)$ (defines space of square integrable functions) can be represented as:

$$f(t) = \sum_{j=1}^{L} \sum_{k=-\infty}^{\infty} d(j,k)\psi(2^{-j}t-k) + \sum_{k=-\infty}^{\infty} a(L,k)\Phi(2^{-L}t-k) \quad (3.18)$$

The function ψ(t) is known as the mother wavelet, while φ(t) is known as the scaling function. The set of functions $\left\{\sqrt{2^{-L}}\Phi(2^{-L}t-k), \sqrt{2^{-j}}\psi(2^{-j}j-k) \mid j \le L \quad j,k,L \in Z\right\}$, where Z is the set of integers, is an orthonormal basis for $L^2(\mathbf{R})$.

The numbers *a(L, k)* are known as the approximation coefficients at scale L, while *d(j,k)* are known as the detail coefficients at scale j. The approximation and detail coefficients can be expressed as:

$$a(L,k) = \frac{1}{\sqrt{2^L}} \int_{-\infty}^{\infty} f(t)\Phi(2^{-L}t-k)dt \qquad \text{a)}$$

$$d(j,k) = \frac{1}{\sqrt{2^j}} \int_{-\infty}^{\infty} f(t)\psi(2^{-j}t-k)dt \qquad \text{b)} \qquad 19)$$

To provide some understanding of the above coefficients consider a projection *fl(t)* of the function *f(t)* that provides the best approximation (in the sense of minimum error energy) to *f(t)* at a scale *l*. This projection can be constructed from the coefficients *a(L,k)*, using the equation:

$$f_l(t) = \sum_{k=-\infty}^{\infty} a(j,k)\psi(2^{-l}t-k) \quad (3.20)$$

As the scale *l* decreases, the approximation becomes finer, converging to *f(t)* as $l \to 0$. The difference between the approximation at scale *l* + 1 and that at *l*, *fl*+1*(t)* - *fl(t)*, is

completely described by the coefficients *d(j, k)* using the equation:

$$f_{l+1}(t) - f_l(t) = \sum_{k=-\infty}^{\infty} d(l,k)\psi(2^{-l}t - k) \quad (3.21)$$

Using these relations, given *a(L, k)* and {*d(j, k)* | j ≤ L}, it is clear that we can build the approximation at any scale. Hence, the wavelet transform breaks the signal up into a coarse approximation *fL(t)* (given *a(L, k)*) and a number of layers of detail {*fj*+1*(t)-fj(t)*| j< L} (given by {*d(j, k)* | j ≤ L}). As each layer of detail is added, the approximation at the next finer scale is achieved.

3.1.3.5 Vanishing Moments

The number of vanishing moments of a wavelet indicates the smoothness of the wavelet function as well as the flatness of the frequency response of the wavelet filters (filters used to compute the DWT) [33].
Typically a wavelet with p vanishing moments satisfies the following equation [32]:

$$\int_{-\infty}^{\infty} t^m \psi(t)dt = 0 \qquad m = 0,..., p-1 \quad (3.22)$$

or equivalently,

$$\sum_j (-1)^k k^m c(k) = 0 \qquad m = 0,..., p-1 \quad (3.23)$$

For the representation of smooth signals, a higher number of vanishing moments leads to a faster decay rate of wavelet coefficients. Thus, wavelets with a high number of vanishing moments lead to a more compact signal representation and are hence useful in coding applications. However, in general, the length of the filters increases with the number of vanishing moments and the complexity of computing the DWT coefficients increases with

the size of the wavelet filters.

3.1.3.6 Subband Coding

The Discrete Wavelet Transform (DWT) coefficients can be computed by using Mallat's Fast Wavelet Transform algorithm. This algorithm is sometimes referred to as the *two-channel sub-band coder* and involves filtering the input signal based on the wavelet function used. To explain the implementation of the Fast Wavelet Transform algorithm consider the following equations:

$$\Phi(t) = \sum_k c(k)\Phi(2t-k) \quad (3.24)$$

$$\psi(t) = \sum_k (-1)^k c(1-k)\Phi(2t-k) \quad (3.25)$$

$$\sum_k c_k c_{k-2m} = 2\delta_{0,m} \quad (3.26)$$

The first equation is known as the *twin-scale relation* (or the dilation equation) and defines the scaling function φ. The next equation expresses the wavelet ψ in terms of the scaling function φ. The third equation is the condition required for the wavelet to be orthogonal to the scaling function and its translates.

The coefficients c(k) or $\{c_0, .., c_{2N-1}\}$ in the above equations represent the impulse response coefficients for a low pass filter of length 2N, with a sum of 1 and a norm of $\frac{1}{\sqrt{2}}$. The high pass filter is obtained from the low pass filter using the relationship $g_k = (-1)^k c(1-k)$, where *k* varies over the range (1-(2N-1)) to 1.

Equation 3.19a shows that the scaling function is essentially a low pass filter and is used to define the approximations. The wavelet function defined by equation 3.19b is a high pass filter and defines the details. Starting with a discrete input signal vector *s*, the first stage of the FWT algorithm decomposes the signal into two sets of coefficients. These are the approximation coefficients cA1 (low frequency information) and the detail coefficients cD1 (high frequency information), as shown in the figure below.

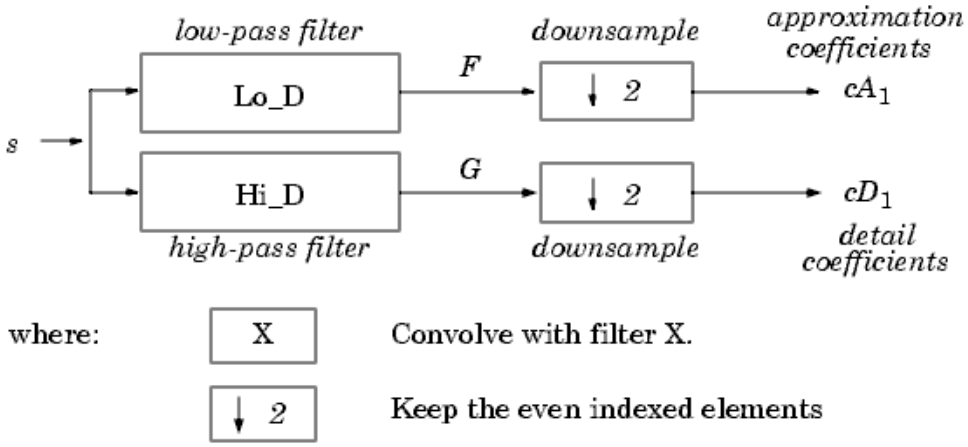


Figure 3.6: Filtering operation of the DWT

The coefficient vectors are obtained by convolving s with the low-pass filter Lo_D for approximation and with the high-pass filter Hi_D for details. This filtering operation is then followed by dyadic decimation or down sampling by a factor of 2.

Mathematically the two-channel filtering of the discrete signal s is represented by the expressions:

$$cA_1 = \sum_k c_k s_{2i-k}, \qquad cD_1 = \sum_k g_k s_{2i-k} \quad (3.27)$$

These equations implement a convolution plus down sampling by a factor 2 and give the forward fast wavelet transform. If the length of each filter is equal to 2N and the length of the original signal s is equal to *n*, then the corresponding lengths of the coefficients cA1 and cD1 are given by the formula:

$$floor\left(\frac{n-i}{2}\right)+N \quad (3.28)$$

This shows that the total length of the wavelet coefficients is always slightly greater than the length of the original signal due to the filtering process used.

3.1.3.7 Multilevel Decomposition

The decomposition process can be iterated, with successive approximations being decomposed in turn, so that one signal is broken down into many lower resolution components. This is called the wavelet decomposition tree.

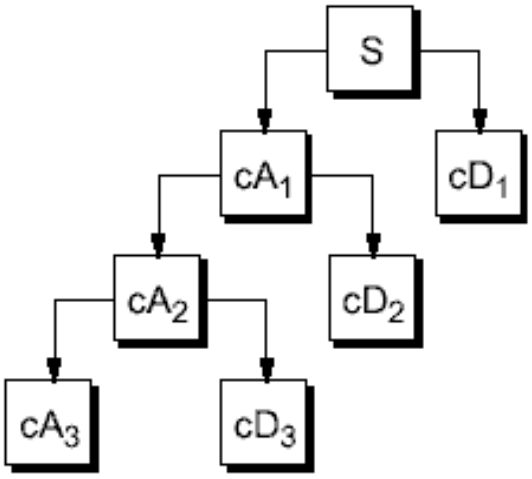


Figure 3.7: Decomposition of DWT coefficients

The wavelet decomposition of the signal s analysed at level j has the following structure [cAj, cDj, ..., cD1]. Looking at a signals wavelet decomposition tree can reveal valuable information. The diagram below shows the wavelet decomposition to level 3 of a sample signal S [31].

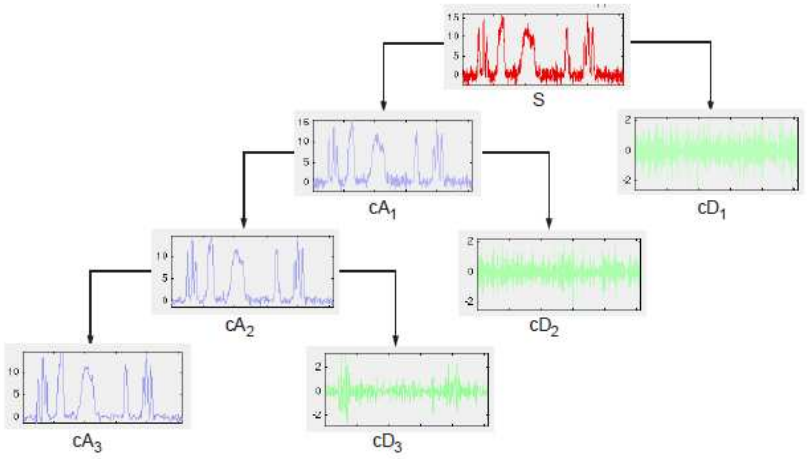


Figure 3.8: Level 3 Decomposition of Sample Signal S [34]

Since the analysis process is iterative, in theory it can be continued indefinitely. In reality, the decomposition can only proceed until the vector consists of a single sample. Normally, however there is little or no advantage gained in decomposing a signal beyond a certain level. The selection of the optimal decomposition level in the hierarchy depends on the nature of the signal being analysed or some other suitable criterion, such as the low-pass filter cut-off [34].

3.1.3.8 Signal Reconstruction

The original signal can be reconstructed or synthesised using the inverse discrete wavelet transform (IDWT). The synthesis starts with the approximation and detail coefficients cA_j and cD_j, and then reconstructs cA_{j-1} by up sampling and filtering with the reconstruction filters.

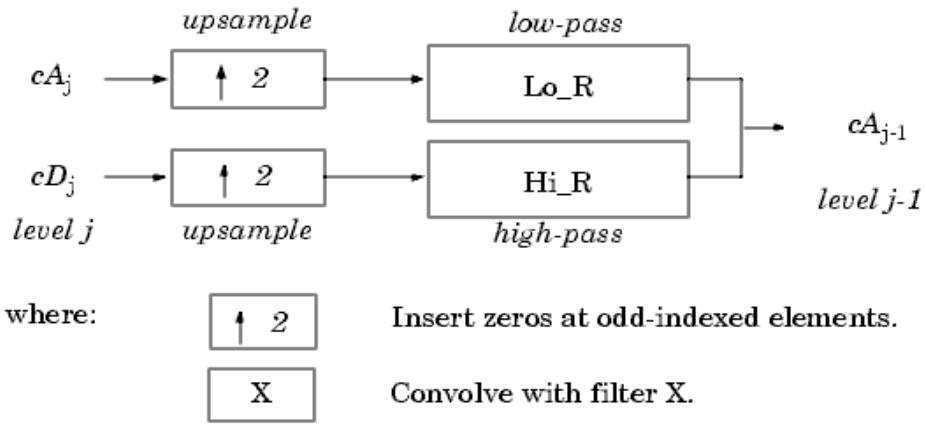


Figure 3.9: Wavelets Reconstruction

The reconstruction filters are designed in such a way to cancel out the effects of aliasing introduced in the wavelet decomposition phase. The reconstruction filters (Lo_R and Hi_R) together with the low and high pass decomposition filters, forms a system known as *quadrature mirror filters* (QMF).

For a multilevel analysis, the reconstruction process can itself be iterated producing successive approximations at finer resolutions and finally synthesising the original signal [34].

3.2 Statistical Analysis

It is important verifying statistical significance in extracted features by HRV signal distinguishing significative changes in data. The use of the word *significance* in statistics is different from the standard one, which suggests that something is important or meaningful. The amount of evidence required to accept that an event is unlikely to have arisen by chance is known as the significance level or critical *p-value*, which is the probability conditional on the null hypothesis of the observed data or more extreme data. If the obtained p-value is small then it can be said either the null hypothesis is false or an unusual event has occurred. It is worth stressing that p-values do not have any repeat sampling interpretation. The significance level is usually denoted by the Greek symbol, α (alpha). In this work level of significance is set to 5% (0.05) [36].

3.2.1 T-Test

The t-test assesses whether the means of two groups are *statistically* different from each other. This analysis is appropriate whenever you want to compare the means of two groups, and especially appropriate as the analysis for the posttest-only two-group randomized experimental design.

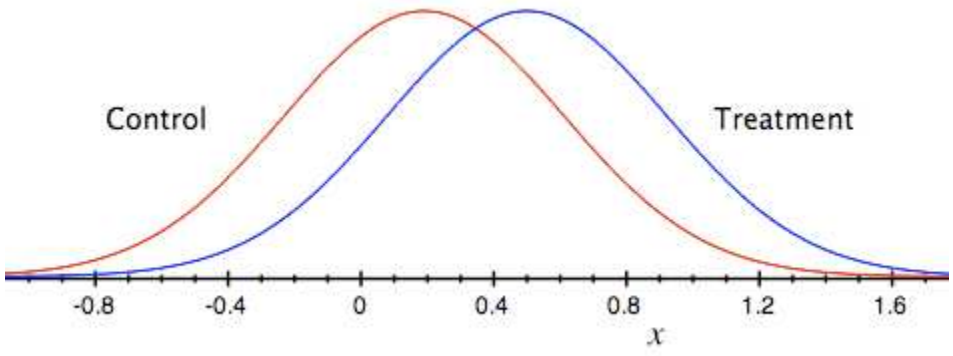


Figure 3.10: Idealized distributions for treated and comparison group posttest values.

Figure 3.10 shows the distributions for the treated (blue) and control (green) groups in a

study. Actually, the figure shows the idealized distribution. The real distribution would usually be with a histogram or bar graph. The figure indicates where the control and treatment group means are located. Considering the three situations shown in figure 3.11, the first thing to notice about the three situations is that the difference between the means is the same in all three, but the three situations don't look the same. The first example shows a case with moderate variability of scores within each group. The second situation shows the high variability case. The third shows the case with low variability.

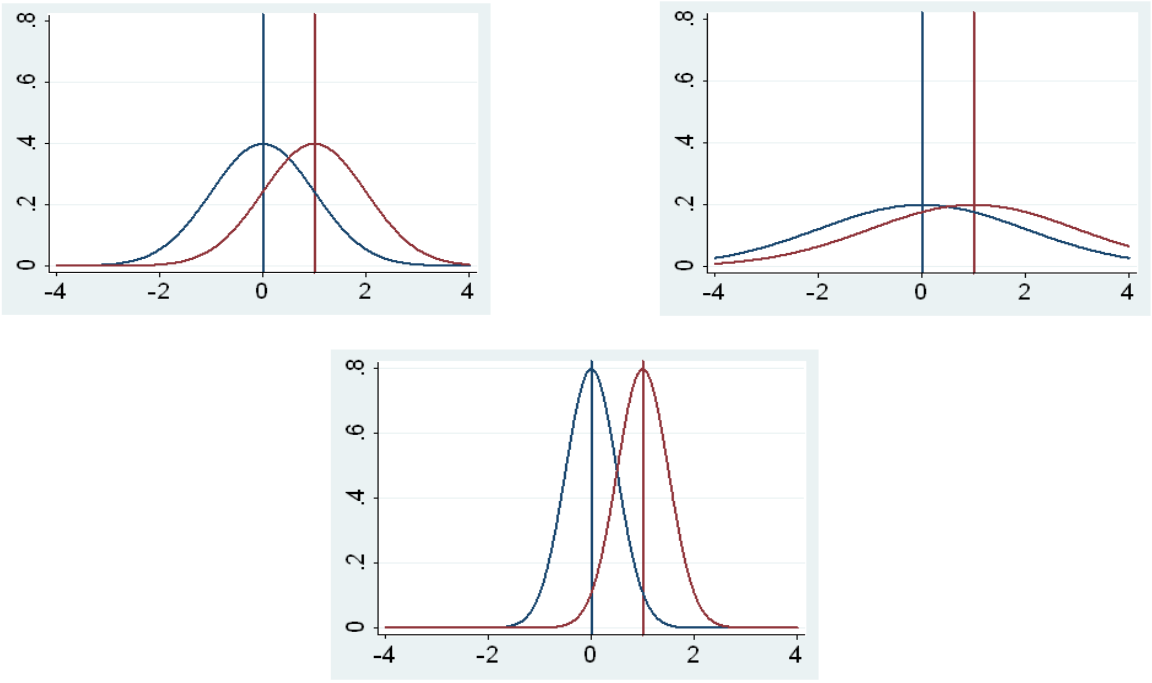


Figure 3.11: Three scenarios for differences between means.

The two groups appear most different or distinct in the bottom or low-variability case because there is relatively little overlap between the two bell-shaped curves. In the high variability case, the group difference appears least striking because the two bell-shaped distributions overlap so much (Figure 3.11). The formula for the t-test is:

$$t = \frac{\overline{X}_T - X_C}{SE\left(\overline{X}_t - \overline{X}_C\right)} \quad (3.29)$$

where the top part of the formula is the difference between the means and the bottom part is called the standard error of the difference. To compute it, the variance for each group is

divided it by the number of people in that group. These two values are added and then are squared root in this way:

$$SE\left(\overline{X}_T - \overline{X}_C\right) = \sqrt{\frac{\text{var}_T}{n_T} + \frac{\text{var}_C}{n_C}} \quad (3.30)$$

The t-value will be positive if the first mean is larger than the second and negative if it is smaller. Once computing the t-value is compared with a table of significance to test whether the ratio is large enough to say that the difference between the groups is not likely to have been a chance finding with alpha level set at 5%. In the t-test, the degrees of freedom are the sum of the persons in both groups minus 2.
Given the alpha level, the degrees of freedom, and the t-value, the t-value is looked in a standard table of significance to determine whether the t-value is large enough to be significant.

3.2.2 Multivariate analysis of variance (MANOVA)

The purpose of a ttest is to assess the likelihood that the means for two groups are sampled from the same sampling distribution of means. The purpose of an ANOVA is to test whether the means for two or more groups are taken from the same sampling distribution. The purpose of MANOVA is to test whether the vectors of means for the two or more groups are sampled from the same sampling distribution. It gives a measure of the overall likelihood of picking two or more random vectors of means out of the same hat. MANOVA is used to explore how independent variables influence some patterning of response on the dependent variables. Here, one literally uses an analogue of contrast codes on the dependent variables to test hypotheses about how the independent variables differentially predict the dependent variables [38].

3.2.3 Kruskal-Wallis Test

The Kruskal–Wallis test is a non-parametric method for testing equality of population medians among groups. It is identical to a one-way analysis of variance with the data replaced by their ranks.

Since it is a non-parametric method, the Kruskal–Wallis test does not assume a normal population, unlike the analogous one-way analysis of variance. However, the test does assume an identically-shaped and scaled distribution for each group, except for any difference in medians. The method is to rank all data from all groups together from 1 to N, ignoring group membership, and to assign any tied values the average of the ranks they would have received had they not been tied. The test statistic is given by:

$$K = (N-1)\frac{\sum_{i=1}^{g} n_i(\bar{r}_i - \bar{r})}{\sum_{i=1}^{g}\sum_{j=1}^{n_i}(r_{ij} - \bar{r})^2} \quad (3.31)$$

where:

- n_i is the number of observations in group i
- r_{ij} is the rank (among all observations) of observation j from group i
- N is the total number of observations across all groups

the p-value is approximated by $\Pr(\chi^2_{g-1} \geq K)$. If some n_i's are small, the probability distribution of K can be quite different from this chi-square distribution. If a table of the chi-square probability distribution is available, the critical value of chi-square, $\chi^2_{\alpha:g-1}$, can be found by entering the table at $g-1$ degrees of freedom and looking under the desired significance or alpha level. The null hypothesis of equal population medians would then be rejected if $K \geq \chi^2_{\alpha:g-1}$. Appropriate multiple comparisons would then be performed on the group medians [37].

3.3 Data Processing

Data transformation refers to the mathematical application that is, each data point z_i is replaced with the transformed value $y_i=f(z_i)$, where f is a function. Transforms are usually applied so that the data appear to more closely meet the assumptions of a statistical inference procedure that is to be applied, or to improve the interpretability or appearance of graphs. The function that is used to transform the data is invertible, and generally is continuous. The reciprocal and some power transformations can be meaningfully applied to data that include both positive and negative values. However when both negative and positive values are observed, it is more common to begin by adding a constant to all values, producing a set of non-negative data to which any power transform can be applied. Classifying data through Neural Network and linear/quadratic discriminant, it is useful to transform a data set to resemble a normal distribution through one of the power transformations. To assess whether normality has been achieved, a graphical approach is usually more informative than a formal statistical test.

3.3.1 Logarithmic Transformation

A logarithmic transformation may be useful when the mean is proportional to the standard deviation. Under such circumstances the logarithmic transformation can be effective in normalizing distributions that have a moderate positive skew. Since a logarithmic transformation makes a more extreme adjustment than a square-root transformation, it can be employed to normalize distributions which have a more severe positive skew. On a logarithmic scale, the distance between adjacent points on the scale will be less than the distance between the corresponding points on the original scale of measurement. As is the case with the square-root transformation, the logarithmic transformation is often useful for normalizing a dependent variable that is a measure of response time.

The logarithmic transformation is obtained through use of the equation $\widetilde{X} = \ln(X)$. Since a logarithm cannot be computed for the value zero, when one or more zeros or positive numbers close to zero are present in a set of data, the following equation is employed:

$\tilde{X} = \ln(X + 1)$. Since a logarithm cannot be computed for a negative number, a constant (the value of which is a positive number which is minimally greater than one unit above the absolute value of the lowest negative number) can be added to all of the values in a set of data, to insure that each value will be a positive number [41].

The logarithm function tends to squeeze together the larger values in data set and stretches out the smaller values. This squeezing and stretching can correct one or more of the following problems with data:

1. Skewed data
2. Outliers
3. Unequal variation

Not all data sets will suffer from these problems. Even if they do, the log transformation is not guaranteed to solve these problems. If data are skewed to the right, a log transformation can sometimes produce a data set that is closer to symmetric.

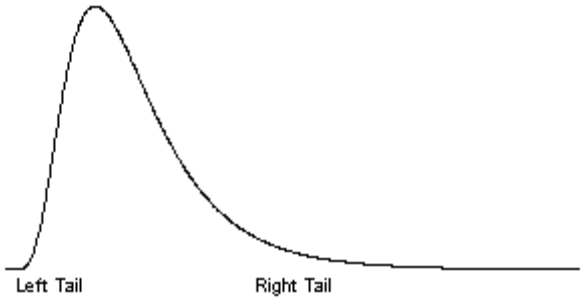


Figure 3.12: Skewed right distribution: the left tail (the smaller values) is tightly packed together and the right tail (the larger values) is widely spread apart.

The logarithm will squeeze the right tail of the distribution and stretch the left tail, which produces a greater degree of symmetry. If the data are symmetric or skewed to the left, a log transformation could actually make things worse. Also, a log transformation is unlikely to be effective if the data has a narrow range (if the largest value is not more than three times bigger than the smallest value) [39].

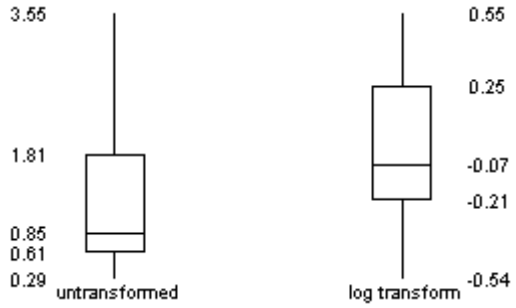


Figure 3.13: Untransformed data shows quite a bit of skewness, while the log transformed data, not perfectly symmetric, does tend to have a better balance between the lower half and the upper half of the distribution.

If data has outliers on the high end, a log transformation can sometimes help. The squeezing of large values might pull that outlier back in closer to the rest of the data. If data has outliers on the low end, the log transformation might actually make the outlier worse, since it stretches small values.

Figure 3.14: The original data has two outliers which are almost 7 standard deviations above the mean. The log transformed data are not perfect, and perhaps there is now an outlier on the low end. The influence of outliers is much less extreme with the log transformed data.

If data has unequal variation, then the some of tests and confidence intervals may be invalid. A log transformation can help with certain types of unequal variation. A common pattern of unequal variation is when the groups with the large means also tend to have large standard deviations. The log transformation will squeeze the groups with the larger standard deviations more than it will squeeze the groups with the smaller standard deviations (Figure 3.15). The log transformation is especially effective when the size of a group's standard deviation is directly proportional to the size of its mean.

Report

DM/DX ratio

Functional alleles	Mean	N	Std. Deviation
No functional alleles	1.272	15	1.036
One or more functional alleles	.013	191	.025
Total	.104	206	.426

Report

log DM/DX ratio

Functional alleles	Mean	N	Std. Deviation
No functional alleles	-.018	15	.335
One or more functional alleles	-2.281	191	.531
Total	-2.116	206	.785

Figure 3.15: The patients with no functional alleles are further from the lower bound and thus have much more room to vary. After a log transformation, the standard deviations are much closer.

A log transformation can sometimes simplify statistical models. Some statistical models are multiplicative: factors influence your outcome measure through multiplication rather than addition. These multiplicative models are easier to work with after a log transformation [40].

3.3.2 Square-root Transformation

A square-root transformation may be useful when the mean is proportional to the variance. Under such circumstances the square-root transformation can be effective in normalizing distributions that have a moderate positive skew, as well as making the treatment variances more homogeneous. This is the case since a square-root scale will reduce the magnitude of difference between the two tails of positively skewed distribution by pulling the right side of the distribution in toward the middle. Consequently a square-root transformation may be able normalize a set of reaction time data. In such a case the square-root transformation can reduce skewness and stabilizes distributional variance. Data taken from a Poisson distribution are sometimes effectively normalized with a square-root transformation. In Poisson distributed data, the mean and variance are proportional.

The square-root transformation is obtained through use the equation $\overline{Y} = \sqrt{X}$, where X is the original score and Y represents the transformed score [41].

3.4 Machine Learning

The purpose of this work is classifying sleep stages in 3 states: WAKE, REM, NREM. Classifiers are feed by the data collected (HRV signal) and obtaining outputs (hypnograms). There are several methods in machine learning theory, which are used in this work, as Neural Network, Linear/Quadratic Discriminant and K-Nearest-Neighbour. Sleep stages are classified by different features, which are extracted by a parametric (RLS algorithm) and wavelet based method (see Chapter 3.1.2 and 3.1.3). Features are selected by sequential forward selection (SFS) algorithm based on Linear Discriminant maximizing Cohen's Kappa index. Understanding these methods needs a general introduction to these techniques.

3.4.1 Linear/Quadratic Discriminant Analysis

Discriminant analysis is a statistical technique to classify objects into mutually exclusive and exhaustive groups based on a set of measurable object's features. Term discriminant analysis comes with many different names for difference field of study. If the number of classes is more than two, it is also sometimes called Multiple Discriminant Analysis (MDA). The purpose of Discriminant Analysis, used in this work, is to classify epochs into one of sleep stages (WAKE, NREM, REM) based on a set of features that describe the stages (e.g. SDNN, HF/TP, LF/TP, pole module HF. movements, etc.). In general, it assigns an epoch to one of a number of predetermined sleep stages based on features characteristics on the epoch. Thus, in discriminant analysis, the dependent variable (Y) is sleep stage and the independent variable (X) is the epoch features that might describe the group. Discriminant assumes that the groups are linearly separable. Linearly separable suggests that the groups can be separated by a linear combination of features that describe the objects. If only two features, the separators between objects group will become lines. If the features are three, the separator is a plane and the number of features (i.e. independent variables) is more than 3, the separators become a hyper-plane.

Using classification criterion to minimize *total error of classification* (TEC), the

proportion of object that misclassifies, must be small as possible. TEC should be thought as the probability that the rule under consideration will misclassify an object. The classification rule is to assign an object to the group with highest conditional probability. This is called Bayes Rule. This rule also minimizes the TEC. Having *K* groups, the Bayes' rule is to assign the object to group *i* where:

$$P(i \mid x) > P(j \mid \bar{x}), \quad \forall j \neq i \quad (3.32)$$

It is important to know the probability $P(i|x)$ that an object is belong to group *i*, given a set of measurement *x*. In practice however, the quantity of $P(i|x)$ is difficult to obtain, but it gets $P(x|i)$, that is the probability of getting a particular set of measurement *x* given that the object comes from group *i*. There is a relationship between the two conditional probabilities that well known as Bayes Theorem:

$$P(i \mid x) = \frac{P(x|i)P(i)}{\sum_{j=1}^{K} P(x|j)P(j)} \quad (3.33)$$

where $P(i)$ is the prior probability of the group *i* known without making any measurement. In practice we can assume the prior probability is equal for all groups or based on the number of sample in each group. In practice, however, to use the Bayes rule directly is unpractical because to obtain $P(x|i)$ need so much data to get the relative frequencies of each groups for each measurement. It is more practical to assume the distribution and get the probability theoretically. Assuming that each group has multivariate-normal distribution (Figure 3.16), described as:

$$P(X = x \mid i = k) = \frac{1}{(2\pi)^{p/2}|C_i|^{1/2}} e^{-\frac{1}{2}(x-\mu_i)^T C_i^{-1}(x-\mu_i)} \quad (3.34)$$

and the denominator is the same for each class, it is possible ignoring it and making the log of the numerator, so the quadratic discriminant function is :

$$f_i(x) = -\frac{1}{2}\log|C_i| - \frac{1}{2}(x-\mu_i)^T C_i^{-1}(x-\mu_i) - \log(p(i)) \quad (3.35)$$

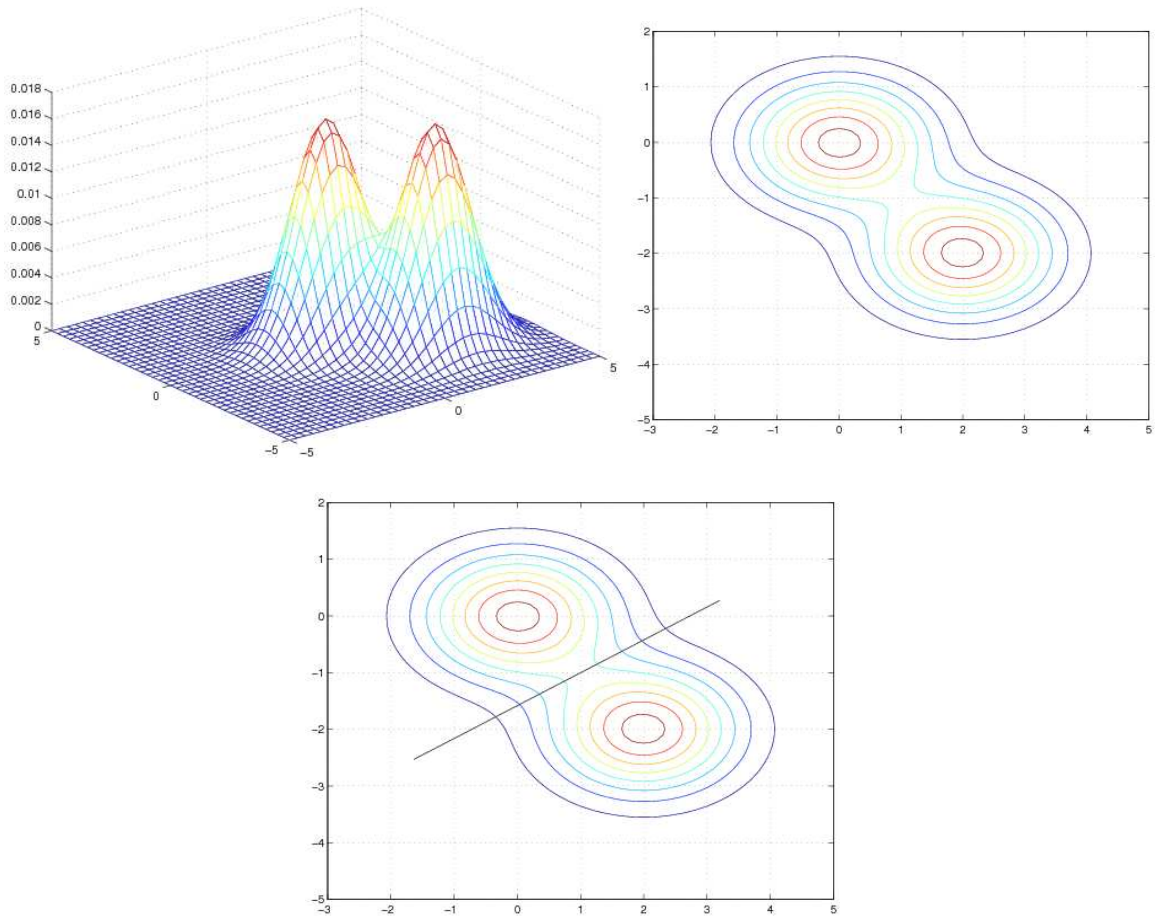


Figure 3.16: The top panel shows the Gaussian distributions that are shifted versions of each other. The bottom panel shows classification boundary obtained by LDA

If it further assumes that the covariance matrices, C_i are equal for each class, then more terms cancel out and linear discriminant function is derived as:

$$f_i = \mu_i C^{-1} x_k^T - \frac{1}{2}\mu_i C^{-1}\mu_i^T + \ln(p(i)) \quad (3.36)$$

where:

- μ_i is vector mean of group *i*
- C_i is covariance marix of group *i*
- x_i is measurement of group *i*

Classification is then performed with these discriminant functions using the following decision rule:

$$G(x) = \arg\max(x_i\, f_i(i)) \quad (3.37)$$

In other words, select the class, *i*, with the highest score for each observation. The second term ($\mu_i C^{-1} \mu_i^T$) is actually called as Mahalanobis Distance, which is distance to measure dissimilarity between several groups.

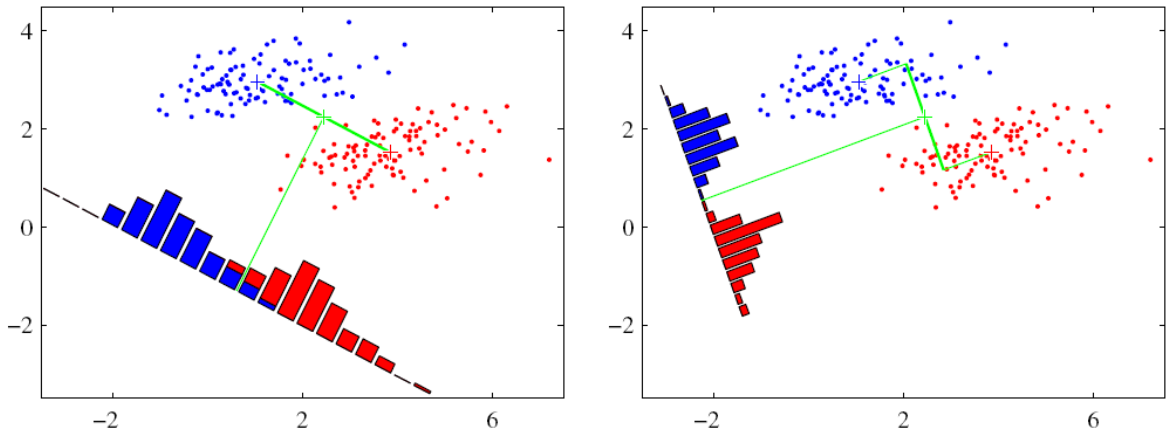


Figure 3.17: The left plot shows samples from two classes along with the histograms Resulting from projection onto the line joining the class means (overlap in the projected space). The right plot shows the corresponding projection based on Mahalanobis distance (better separation).

3.4.2 Neural Network

In subparagraph 3.4.1 models for regression and classification, that comprised linear combinations of fixed basis functions, are described. These models have useful analytical

and computational properties but that their practical applicability was limited by the curse of dimensionality. In order to apply such models to largescale problems, it is necessary to adapt the basis functions to the data, in other words to use parametric forms for the basis functions in which the parameter values are adapted during training. The most successful model of this type in the context of pattern recognition is the feed-forward neural network, also known as the *multilayer perceptron*, discussed in this chapter. In fact, 'multilayer perceptron' is really a misnomer, because the model comprises multiple layers of logistic regression models (with continuous nonlinearities) rather than multiple perceptrons (with discontinuous nonlinearities). For many applications, the resulting model can be significantly more compact, and hence faster to evaluate, than a support vector machine having the same generalization performance. The price to be paid for this compactness, as with the relevance vector machine, is that the likelihood function, which forms the basis for network training, is no longer a convex function of the model parameters. In practice, however, it is often worth investing substantial computational resources during the training phase in order to obtain a compact model that is fast at processing new data ([69], [70], [71]).

3.4.2.1 Feed-forward Network Functions

Neural networks use basis functions that follow the same form as:

$$z_k = f\left(\sum_{j=1}^{M} w_j \phi_j(x) \right) \quad (3.38)$$

so that each basis function is itself a nonlinear function of a linear combination of the inputs, where the coefficients in the linear combination are adaptive parameters. This leads to the basic neural network model, which can be described a series of functional transformations. First, M linear combinations of the input variables $x_1, \ldots, x_D$ in the form are constructed as:

$$net_j = \sum_{i=1}^{d} w_{ji} x_i + w_{j0} = \sum_{i=0}^{d} w_{ji} x_i \equiv w_j^T x \quad (3.39)$$

where the subscript i indexes units on the input layer, j for the hidden; w_{ji} denotes the input-to-hidden layer weights at the hidden unit j. Each hidden unit emits an output that is a nonlinear function of its activation, $f(net)$:

$$y_j = f\left(net(j)\right) \quad (3.40)$$

Each output unit similarly computes its net activation based on the hidden unit signals as:

$$net_k = \sum_{j=1}^{n} y_j w_{kj} + w_{k0} = \sum_{j=0}^{n} y_j w_{kj} \equiv w_k^T y \quad (3.41)$$

where the subscript k indexes units in the output layer and n_H denotes the number of hidden units. The bias unit is treated as equivalent to one of the hidden units whose output is always $y0 = 1$. Each output unit then computes the nonlinear function of its *net*, emitting

$$z_k = f\left(net_k\right) \quad (3.42a)$$

Because binary outputs are in this work, hidden and output activation function is been choosen as sigmoid:

$$f\left(net_k\right) = \frac{1}{1+e^{-net_k}} \quad (3.42b)$$

For classification, it defines c output units, one for each of the categories, and the signal from each output unit is the discriminant function z_k*:*

$$z_k = f\left(\sum_{j=1}^{n} w_{kj} f\left(\sum_{i=1}^{d} w_{ji} x_i + w_{j0}\right) + w_{k0}\right) \quad (3.43)$$

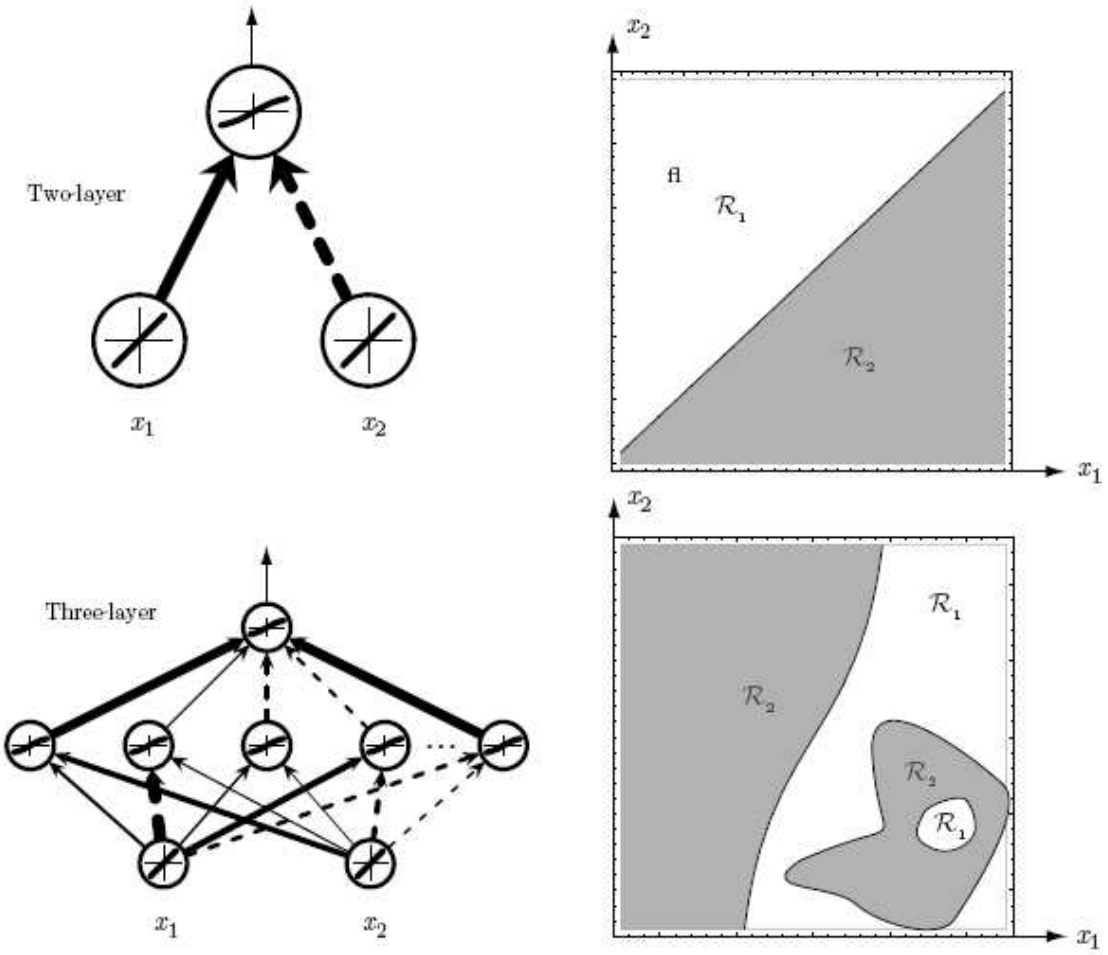


Figure 3.18: Whereas a two-layer network classifier can only implement a linear decision boundary, given an adequate number of hidden units, networks can implement arbitrary decision boundaries.

The neural network model comprises two stages of processing, each of which resembles the perceptron model, and for this reason the neural network is also known as the *multilayer perceptron*, or MLP. A key difference compared to the perceptron, however, is that the neural network uses continuous sigmoidal nonlinearities in the hidden units, whereas the perceptron uses step-function nonlinearities. This means that the neural network function is differentiable with respect to the network parameters, and this property will play a central role in network training [75].

3.4.2.2 Network Training

The basic approach in learning is to start with an untrained network, present an input training pattern and determine the output. The error or criterion function is some scalar function of the weights that is minimized when the network outputs match the desired outputs. The weights are adjusted to reduce this measure of error.

$$J(w)=\frac{1}{2}\sum_{k=1}^{c}(t_k-z_k)^2=\frac{1}{2}(T-Z) \quad (3.44)$$

where T and Z are target and the network output vectors of length *c*; w represents all the weigths in the network. The backpropagation learning rule is based in gradient descent. The weights are initialized with random values, and are changed in a direction that will reduce the error:

$$\Delta w=-\eta\frac{\partial J}{\partial w} \quad (3.45)$$

where η is the *learning rate*.
This iterative algorithm takes a weigth vector at iteration *m* and updates it as:

$$w(m+1)=W(m)+\Delta w(m) \quad (3.46)$$

Because the error is not explicitly dependent upon w_{jk}, defining the sensitivity of unit *k* as:

$$\delta_k\equiv-\frac{\partial J}{\partial net_k} \quad (3.47)$$

which describes how the overall error changes with the unit's activation. In the output layer, the sensitivity for an output unit is defined as:

$$\delta_k=(t_k-z_k)f'(net_k) \quad (3.48)$$

and for a hidden unit as:

$$\delta_k \equiv f'(net_k)\sum_{k=1}^{c} w_{kj} d_k \quad (3.49)$$

The sensitivity at a hidden unit is simply the sum of the individual sensitivities at the output units weighted by the hidden-to-output weights w_{jk}, all multiplied by $f'(net_j)$. Thus the learning rule for the input-to-hidden weights is:

$$\Delta w_{ji} = \eta x_i \delta_j = \eta x_i f'(net_j)\sum_{k=1}^{c} w_{kj}\delta_k \quad (3.50)$$

Equations 3.48 & 3.50, give the backpropagation algorithm, so-called because during training an error (actually, the sensitivities δ_k) is propagated from the output layer *back* to the hidden layer in order to perform the learning of the input-to-hidden weights by Equation 3.50. At base then, backpropagation is "just" gradient descent in layered models where the chain rule through continuous functions allows the computation of derivatives of the criterion function with respect to all model parameters (i.e., weights).

One of the problems that occur during neural network training is called overfitting. The error on the training set is driven to a very small value, but when new data is presented to the network the error is large. The network has memorized the training examples, but it has not learned to generalize to new situations [76].

3.4.2.3 Levenberg-Marquardt

Like the quasi-Newton methods, the Levenberg-Marquardt algorithm was designed to approach second-order training speed without having to compute the Hessian matrix. When the performance function has the form of a sum of squares (as is typical in training feedforward networks), then the Hessian matrix can be approximated as

$$H = J^T J \quad (3.51)$$

and the gradient can be computed as

$$g = J^T e \quad (3.52)$$

where J is the Jacobian matrix that contains first derivatives of the network errors with respect to the weights and biases, and e is a vector of network errors. The Jacobian matrix can be computed through a standard backpropagation technique [72] that is much less complex than computing the Hessian matrix. The Levenberg-Marquardt algorithm uses this approximation to the Hessian matrix in the following Newton-like update:

$$x_{k+1} = x_k - [J^T J + \mu I]^{-1} J^T e \quad (3.53)$$

When the scalar μ is zero, this is just Newton's method, using the approximate Hessian matrix. When μ is large, this becomes gradient descent with a small step size. Newton's method is faster and more accurate near an error minimum, so the aim is to shift toward Newton's method as quickly as possible. Thus, μ is decreased after each successful step (reduction in performance function) and is increased only when a tentative step would increase the performance function. In this way, the performance function is always reduced at each iteration of the algorithm [75].

3.4.2.4 Early Stopping

The default method for improving generalization is called *early stopping.* In this technique the available data is divided into three subsets. The first subset is the training set, which is used for computing the gradient and updating the network weights and biases. The second subset is the validation set. The error on the validation set is monitored during the training process. The validation error normally decreases during the initial phase of training, as does the training set error. However, when the network begins to overfit the data, the error on the validation set typically begins to rise (Figure 3.19). When the validation error

increases for a specified number of iterations, the training is stopped, and the weights and biases at the minimum of the validation error are returned [73].

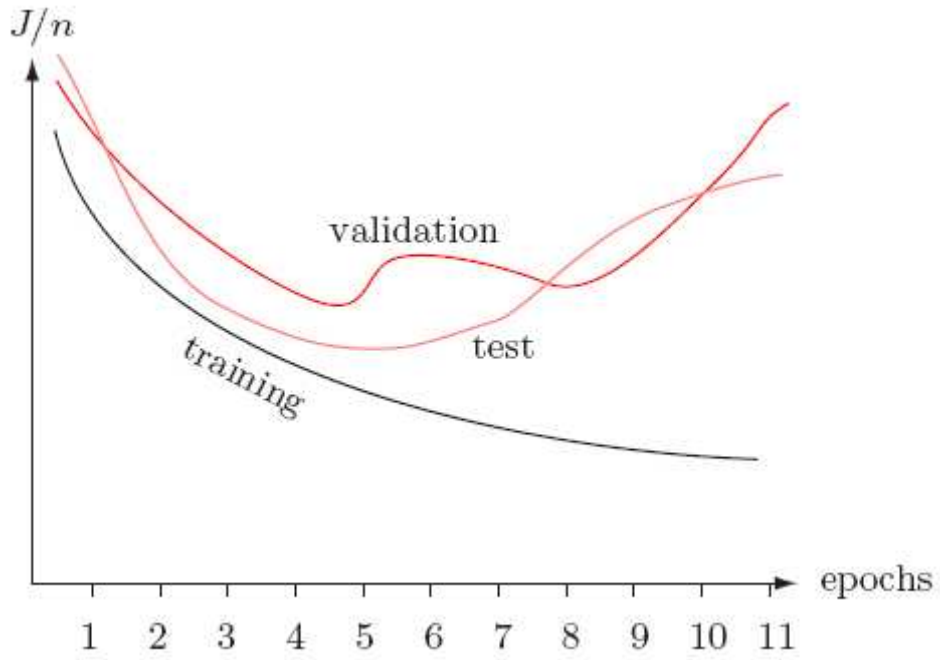


Figure 3.19: The validation error and the test error per pattern are virtually always higher than the training error. Training is stopped at the minimum of the validation set.

The test set error is not used during training, but it is used to compare different models. It is also useful to plot the test set error during the training process. If the error in the test set reaches a minimum at a significantly different iteration number than the validation set error, this might indicate a poor division of the data set [75].

3.4.2.5 Number of hidden units

While the number of input units and output units are dictated by the dimensionality of the input vectors (features) and the number of categories (sleep stages), respectively, the number of hidden units is not simply related to such obvious properties of the classification problem. The number of hidden units, *nH*, governs the complexity of the decision boundary. If the patterns are well separated or linearly separable, then few hidden units are needed; conversely, if the patterns are drawn from complicated densities that are highly interspersed, then more hiddens are needed. Thus without further information there is no

foolproof method for setting the number of hidden units before training. Figure 3.20 shows the training and test error on a three-sleep stages classification problem for networks that differ solely in their number of hidden units. For large *nH*, the training error can become small because such networks have high expressive power and become tuned to the particular training data. At the other extreme of too few hidden units, the net does not have enough free parameters to fit the training data well, and again the test error is high. We seek some intermediate number of hidden units that will give low test error [76].

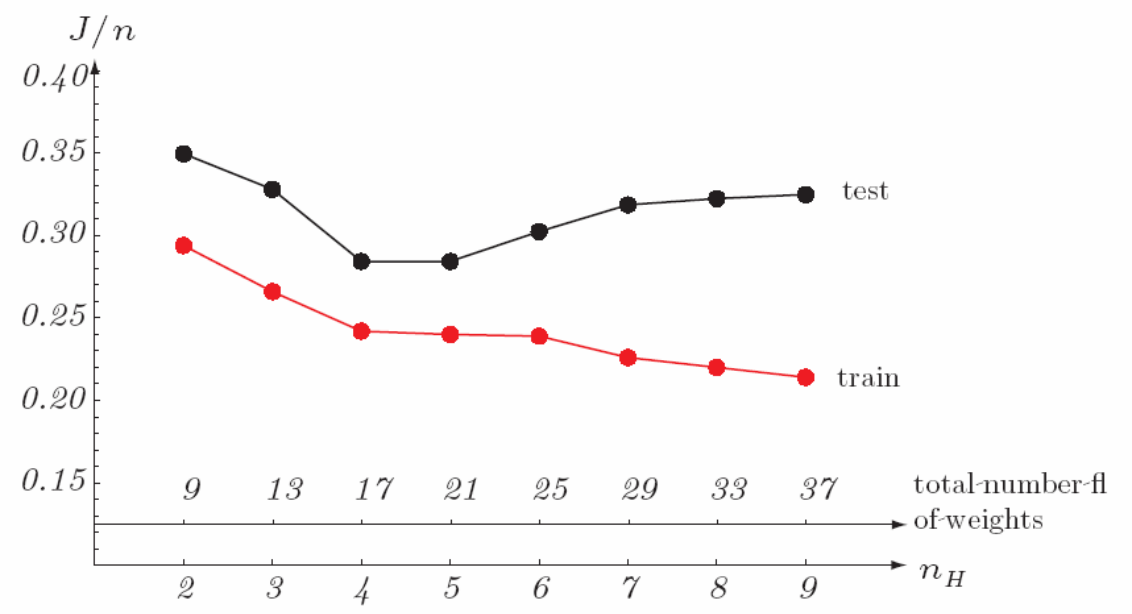


Figure 3.20. Training and test classification error respect to the number of hidden layer neurons.

3.4.2.6 Error surfaces

Since backpropagation is based on gradient descent in a criterion function, it is useful studying the function *J(w)*. Of course, such an error surface depends upon the training and classification task; nevertheless there are some general properties of error surfaces that seem to hold over a broad range of real-world pattern recognition problems. If many local minima plague the error landscape, then it is unlikely that the network will find the global minimum. Many times it is useful starting with difference random weight to start training in different points and looking for global minima.

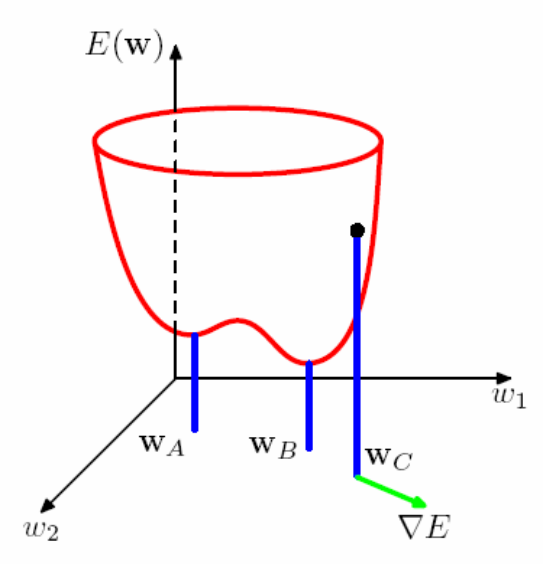


Figure 3.21: Geometrical view of the error function E(w) as a surface sitting over weight space. Point w_A is a local minimum and w_B is the global minimum. At any point w_C, the local gradient of the error surface is given by the vector ∇E .

3.4.3 *K*-Nearest-Neighbor

K-Nearest-Neighbor (*k*-NN) method assumes all instances correspond to points in the *n*-dimensional space. The nearest neighbors of an instance are defined in terms of the standard Euclidean distance. The problem of classification is modelling a discrete-valued target function F: $\Re^n \rightarrow V$, where V is the finite set $\{v_1,...v_s\}$.

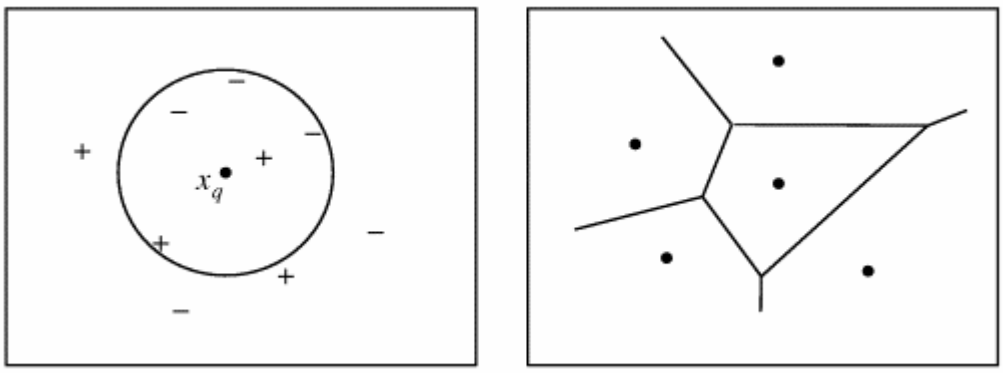


Figure 3.22: K-Nearest Neighbor method. In the left panel, the new instance x_q is classified to "+" or "–". In the right panel, Voronoi diagram is represented

In Figure 3.22, instance are points in 2-dimensional space, the number of classes s = 2, so the output is boolean (denoted as "+" or "–"). New instance x_q (called also a *query point)* is classified with respect to proximity of nearest training instances. A 1-NN method, which uses only 1 training instance (the closest to the query), and x_q will be classified to "+" (since the nearest training instance belongs to class "+"). A 5-NN method accepts 5 training instances and with x_q will be classified to "–" (since among 5 instances there are 2 "+" and 3 "–").The essence of the *K*-NN algorithm is to classify a new x_q finding the most common value of the nearest training instances. The right panel of figure 3.22 shows the Voronoi diagram, a decision surface induced by the 1-NN algorithm for a typical set of training instances. The convex polygon surrounding each training instance indicates the region of instance space closest to that point.

An arbitrary instance x is defined as $\{a_1(x) \ldots a_n(x)\}$, where $a_r(x)$ denotes the value of the r-th attribute of instance x. The distance between two instances x_i and x_j is defined as:

$$d(x_i, x_j) = \sqrt{a_r(x_i)^2 - a_r(x_j)^2} \quad (3.54)$$

One attractive aspect of the nearest-neighbour decision rule is that it is often possible to reduce the size of the training set S without changing the decision boundary, considering the Voronoi diagram of S. A refinement of the *K*-NN classification algorithm is to weigh the contribution of each of the *K* neighbors according to their distance to the query point x_q, giving greater weight w_i to closer neighbors. This can be accomplished by replacing the final line in the algorithm by:

$$F(x_q) = \arg\max\left(\sum_{i=1}^{k} w_i \delta(v, f(x_i))\right) (3.55)$$

where the weight is:

$$w_i = \frac{1}{d(x_q, x_i)^2} (3.56)$$

(in case x_q exactly matches one of x_i, so that the denominator becomes zero, we assign $F(x_q)$ to be $f(x_i)$ in this case. For the version of K-NN for real-valued output the final line of the algorithm will be:

$$F(x_q) = \frac{\sum_{i=1}^{k} w_i f(x_i)}{\sum_{i=1}^{k} w_i} \quad (3.57)$$

If weighting is used, it makes sense to use all training examples, not just K – the algorithm then becomes a *global* one, since all training instances are used. The only disadvantage is that the algorithm will run more slowly [75].

3.5 Statistical performances

As a measure of system performance, accuracy and Cohen's Kappa Coefficient have been used which are measure of interrater agreement, where the two raters are the expert sleep technician (who scored the polysomnograph recordings) and the automated sleep staging system. Accuracy is the proportion of true results (both true positive and true negative) in the population [81]. Table 3.1 shows confusion matrix for calculating accuracy and Cohen's Kappa Coefficient for 2 classes:

		Gold Standard	
		True	False
Test outcome	Positive	**T**rue **P**ositive	**F**alse **P**ositive
	Negative	**F**alse **N**egative	**T**rue **N**egative

Table 3.1: Confusion Matrix

where:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (3.58)$$

Cohen's kappa measures the agreement between two raters who each classify N items into C mutually exclusive classes.

$$k = \frac{p_0 - p_e}{1 - p_e} \quad (3.59)$$

where p_0 is the relative observed agreement among raters, and p_e is the hypothetical probability of chance agreement, using the observed data to calculate the probabilities of each observer randomly saying each category. If the raters are in complete agreement then $\kappa = 1$. If there is no agreement among the raters (other than what would be expected by chance) then $\kappa \leq 0$ [82].

The need for such a measure is evident when we consider the relative proportions of the sleep stages (WAKE, NREM, REM) whose ratios are approximately 25% for WAKE, 65% for NREM, 10% for REM. Therefore, with complete ignorance all stages could be scored as NREM and achieve 65% accuracy, which may appear to be quite a reasonable performance. However, in this instance k=0, which is a better measure of performance.

Chapter 4

Methodology of Protocol

"If a man will begin with certainties, he shall end in doubts; but if he will be content to begin with doubts he shall end in certainties."

Francis Bacon

Sleep recordings were performed on 11 adults (age 20-54 years) at the sleep laboratory of Finnish Institute of Occupational Health (FIOH). Each subject participated with two recordings and these were obtained after baseline night, once during daytime sleep after a night shift of work and once during nighttime sleep. Signals were scored using standard R&K criteria on EEG, EOG and EMG. The sleep scoring, based on standard polysomnographic recordings, was done by expert personnel. R-R intervals (RRI) were computed from the standard ECG signal with the Somnologica software. In addition, the multichannel BCG was recorded with the bed sensor using multiple Emfit electrodes. Both the heart beat interval (HBI), with coverage of 88%, and movement activity were extracted from the bed sensor signals.

FIOH provided:

- Heart beat intervals sampled at 10 Hz
- Respiration signal sampled at 10 Hz
- Movement activity sampled at 10 Hz
- Polysomnographic recordings sampled at 1 Hz

The sleep time was divided in epochs of 30 seconds each. In sleep medicine, a single epoch can be scored as stage 1, 2, 3, 4, REM and Wake. The resulting data were coded as follows:

- 0 no scoring
- -1 movement time

- -2 WAKE
- -3 REM
- -4 Stage 1
- -5 Stage 2
- -6 Stage 3
- -7 Stage 4

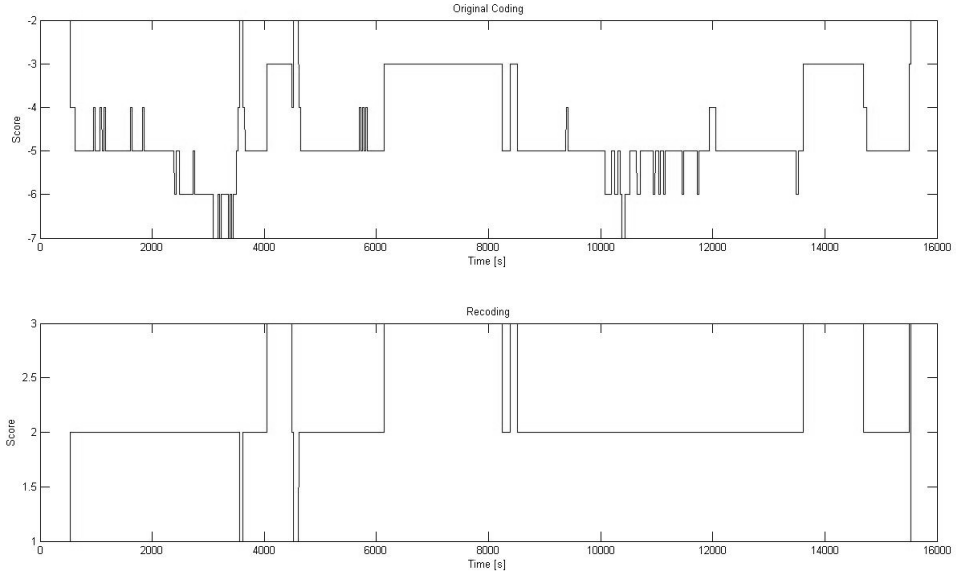


Figure 4.1: The figure shows sleep stage coding performed by FIOH on the top and particular the coding used in this work in the bottom.

In this work, the stages from 1 to 4 were merged in one stage named NREM (Non REM) and stages with the code 0 (*no scoring*, corresponding to light on and subject out of the bed) and the code -1 (*movement time*, subject sat on the bed side), were deleted. So, the sleep hypnogram was re-coded (Figure 4.1) as:

- 1 as WAKE stage
- 2 as NREM stage
- 3 as REM stage

4.1 Data pre-processing

The heart beat intervals, the movement activity and the respiration signals were pre-processed and synchronized with sleep scoring through signal time. According to the bed sensor acquisition, there are some outlier data. Particularly, RR and respiration signals were filtered by ten-degree and sixty-degree median filter respectively. Figure 4.1 shows an example of RR and respiratory signal filtered with a median filter.

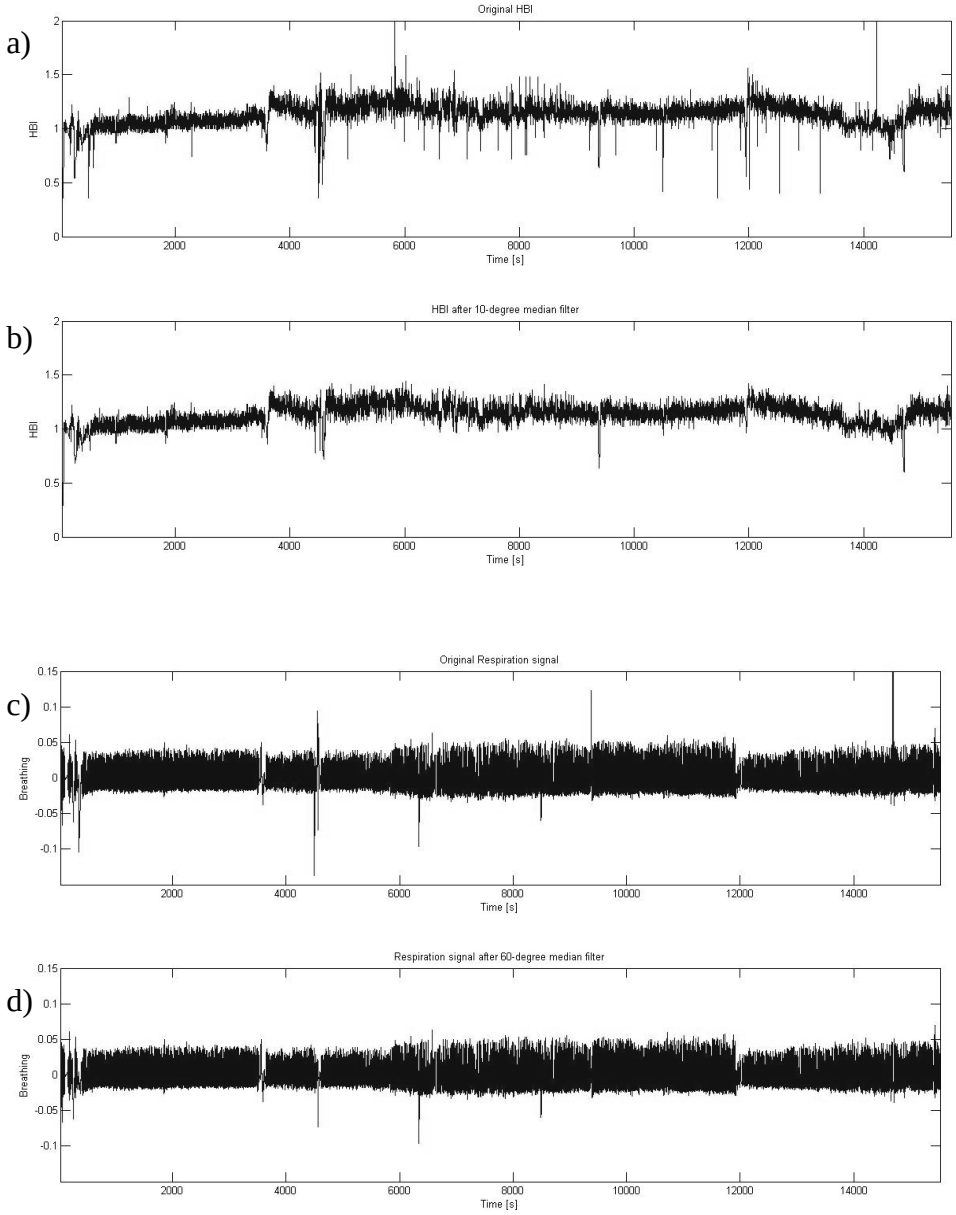


Figure 4.2: HBI and breathing signals of first patient. a), b) effect of 10-degree median filter on HBI. c), d) respiratory outlier data were deleted by 60-degree median filter.

4.2 Features extraction with RLS method

Since HBI as well as RRI become nonstationary for the long-term sleep recording, two time-frequency methods, such as the recursive least square filter (RLS) and the wavelet, result to be the best choices in order to evaluate power spectral density (PSD). This chapter analyzes only the RLS method. According Juha M. Kortelainen et al. [74], the respiration and the RR signals were filtered by high-pass filter, and a forgetting factor that gives more weight to the most recent samples was used. The heart rate variability (HRV) and the respiratory spectrum were computed starting from the obtained autoregressive parameters on a beat-by-beat basis.

A forgetting factor of 0.985 was used for all sleep recordings, while an eight and three order model were used for HRV and respiratory signals, respectively.

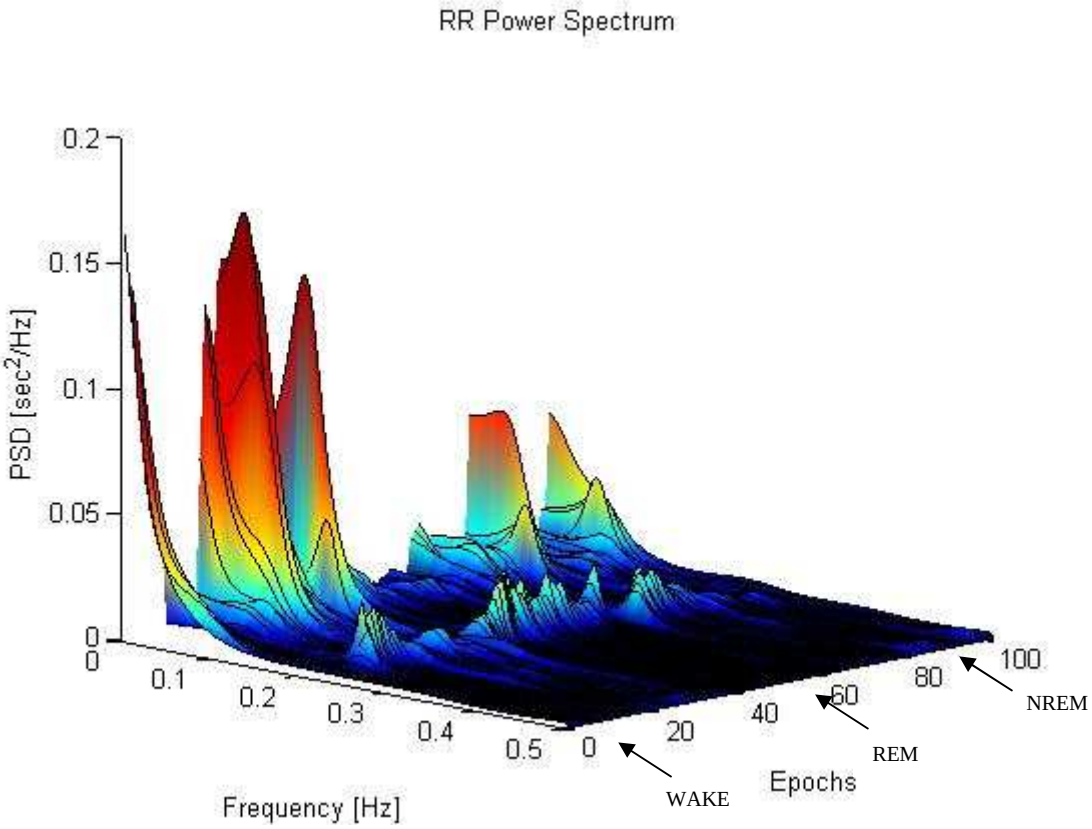


Figure 4.3a: PSD computed by RLS for HBI signal of 1st subject

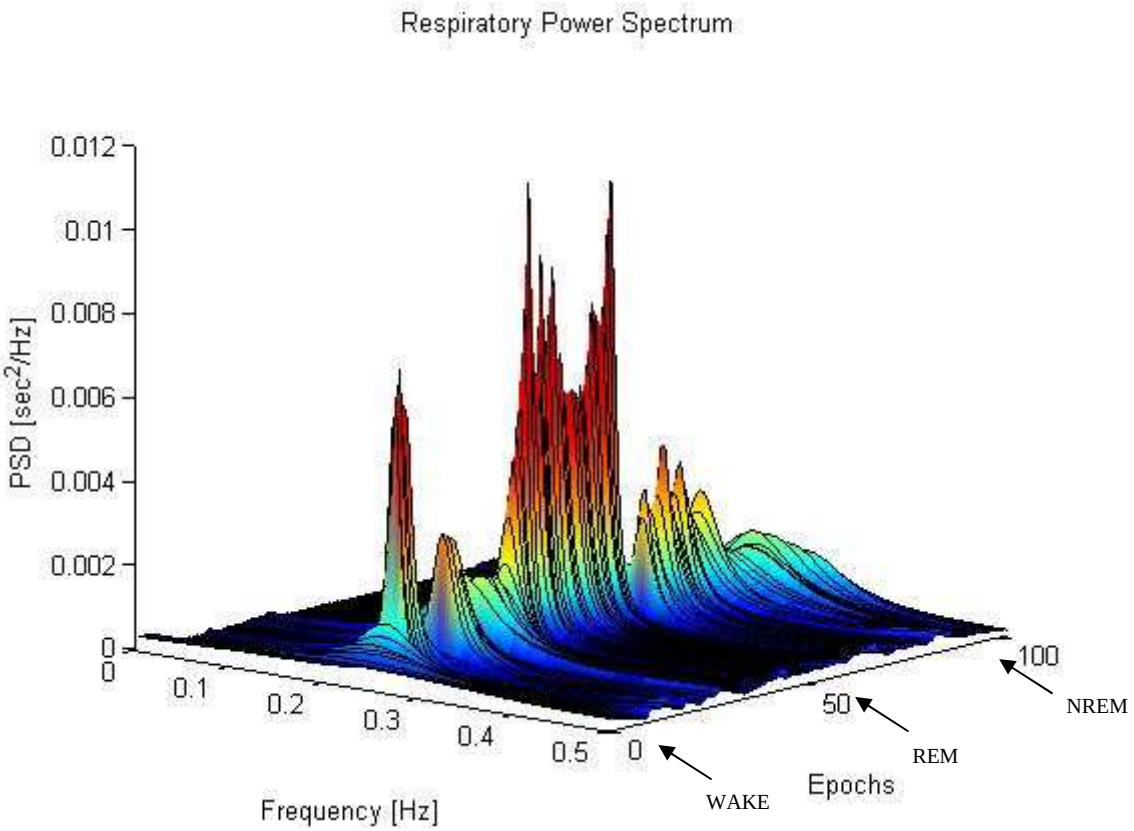


Figure 4.3b: Respiratory PSD computed by RLS for 1st subject

Using the time variant HRV and respiratory spectrum the following beat-by-beat features were obtained:

- Mean on 150 seconds of RR values
- Standard Deviation on 150 seconds of RR values
- Very low frequency power - VLF (0.003 Hz - 0.02 Hz)
- Low frequency power - LF (0.02 Hz - 0.15 Hz)
- High frequency power - HF (0.15 Hz - 0.5 Hz)
- HF and VLF power ratio
- LF and VLF power ratio
- LF and HF power ratio
- Cross-spectrum module of HRV and respiratory signals in the same bands

LF and HF indices are related to the sympathovagal balance and present characteristic values at the different sleep stages [11]. However, these sympathovagal values may be different among subjects due to the biovariability. This inter-subject variability has to be eliminated in order to obtain values that can be comparable between subjects and to enable

the sleep stage classification. This normalization process eliminates some inconsistencies and assures that the noise produced by the bio-diversity of subjects is reduced or damped. The HRV spectral parameters (HF, LF, and VLF) were normalized beat-by-beat as the percentage power with respect to the total power (TP).

These features were divided into epochs of 30 seconds. Each epoch contains the mean value of every single feature. Figure 4.4 and equation 4.1 illustrate how the mean and the standard deviation, for each epoch, were obtained by averaging 150 seconds (from -60 to 60 seconds respect to the central epoch). For each recording, due to the RLS parameters (forgetting factor and model order), the model reaches stability after 80 beats, the first three epochs were deleted as well as the last three epochs were also deleted due to the noise..

$$Mean_{150s} = \frac{1}{150}\sum_{i=-60}^{90} x(i)$$
$$Std\_Dev_{150s} = \sqrt{\frac{1}{150}\sum_{i=-60}^{90}\left(x(i) - Mean_{150x}\right)^2} \quad (4.1)$$

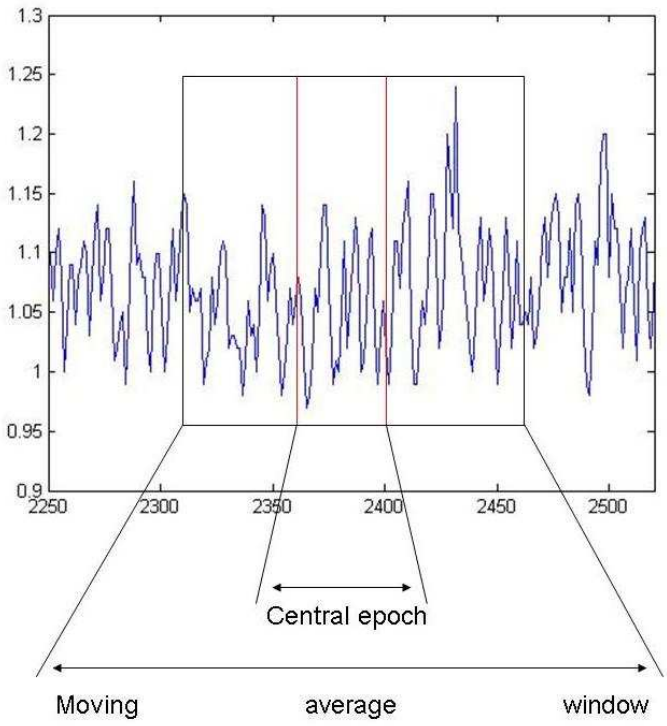


Figure 4.4: A moving window on 150 sec was applied to signals to computing mean and standard deviation, assigning values obtained to central epoch.

4.3 Features extraction from wavelet coefficients

In this chapter, a description the wavelet to HRV is presented. First, the wavelet family has to be chosen. In this work, the Daubechies (figure 4.5 presents 4-order Daubchies mother wavelet) is used in order to compute a 6-level decomposition.

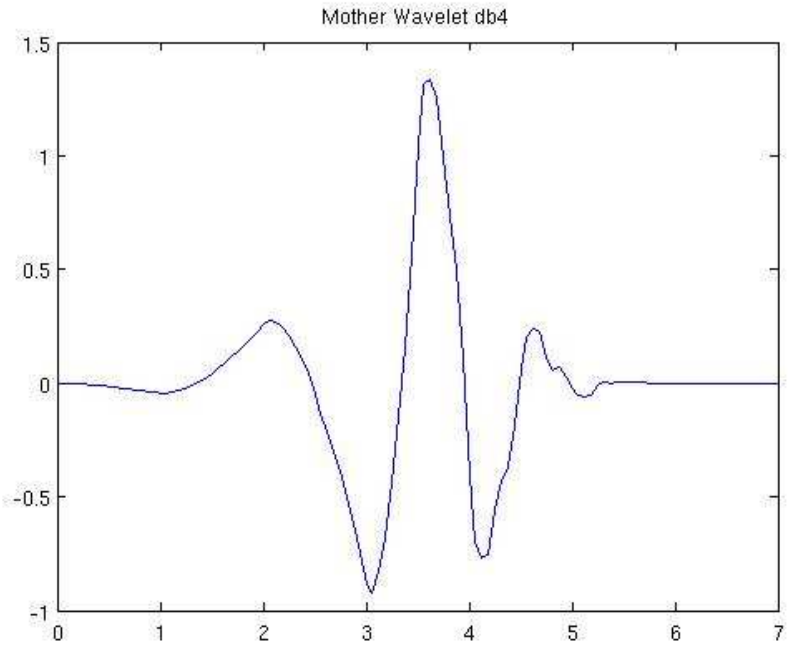


Figure 4.5: Four order Daubechies mother wavelet.

Due to the 1 Hz sampling frequency and a limited range of power (0.003 Hz- 0.5 Hz), a 4 order wavelet with an associated pseudo-frequency has to be used, as shown in the following table 4.1:

Level	Scale	Frequency (Hz)
1	2	0,3571
2	4	0,1786
3	8	0,0893
4	16	0,0446
5	32	0,0223
6	64	0,0112

Table 4.1: Pseudo-frequency associated to each level decomposition

The first level is not considered in order to de-noise signal, as the low-pass filter used in RLS. The scalogram computed is shown in Figure 4.6.

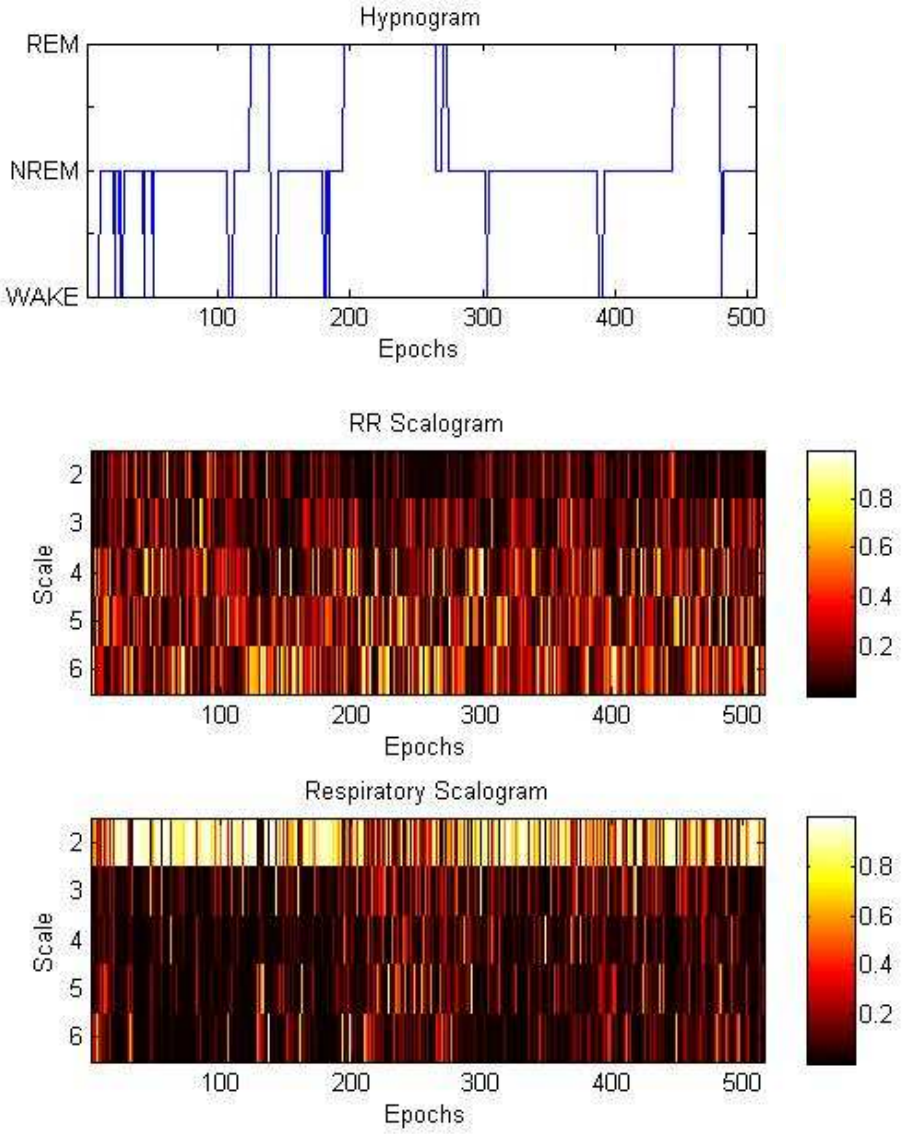


Figure 4.6: Scalogram computed of 1st subject with Daubechies family wavelet of 4 order for both signals. Levels are only used from 2 to 6.The hypnogram relates scalograms with sleep stages

The result consists of 29 beat-by-beat features obtained by mixing the calculated time variant HRV and the respiratory details:

- Mean on 150 seconds of RR values
- Standard Deviation on 150 seconds of RR values
- Detail 1, 2, 3, 4 power
- Detail 1 and detail 2 ratio

- Detail 1 and detail 3 ratio
- Detail 1 and detail 4 ratio
- Detail 2 and detail 3 ratio
- Detail 2 and detail 4 ratio
- Detail 3 and detail 4 ratio
- Cross-spectrum module of HRV and respiratory signals in the same detail

These features were divided into epochs of 30 seconds. Each epoch contains the mean value of every single feature. Details were normalized beat-by-beat as the percentage power with respect to the sum of each detail eliminating inter-subject variability. Obtained values are comparable between subjects and to enable the sleep stage classification.

4.4 Sequential Feature Selection Algorithms

The sequential feature selection (SFS) algorithms search in a sequential deterministic manner for the sub-optimal best feature subset . SFS starts from an empty feature set, and in each iteration generates new subsets by adding a feature selected by some evaluation function, in the present work the Cohen's kappa index (for details see Chapter 3.5) between expert scoring and the automatic one. It evaluates each subset using leave-one-out (LOO), and in each iteration, selects the subset having the highest mean kappa index value on every patient. It continues until mean kappa index values are computed for each adding feature. The kappa index computing method is based on the linear (LD) or the quadratic (QD) discriminant classifier. The first step is repeated iteratively adding a new feature each time; then the new features are the one which maximizes the kappa index value when combined with the other features chosen in the previous steps. Each step gives the sub-optimal combination of features with different length (from 1 to the number of features). When all the possible combinations are computed, the algorithm itself selects the one which has the highest kappa index value.

Pseudo-code for the sequential forward algorithm is given bellow:

1. Initialize the feature subset

$F_0=\{\varnothing\}$; $F_{evaluation}=0$; j=1;

2. Add one feature *f* at time for all features

$F_{evaluation} = F_{evaluation} + f$;

3. For all subjects *s* compute kappa index (LOO method)

4. Mean kappa index on subjects

kappa index(j,s)=mean(kappa index,2);

5. Add a feature based on performance of mean kappa index results

$$x^{+} = \arg\max_{x \notin F_j}[J(F_j + x)]$$

j=j+1;

6. Update the feature subset

$$F_{j+1} = F_j + x$$

If j=length size of features then

END

Else

go to step 2

7. Choose the feature subset length maximing mean kappa index

This algorithm is repeated for each sleep stage (WAKE, REM, NREM) considering only RR extracted features, only respiratory ones or both.

The following figures (from Figure 4.7 to Figure 4.18) show mean kappa index value evolution for all sleep stages with both classifiers used, LD and QD.

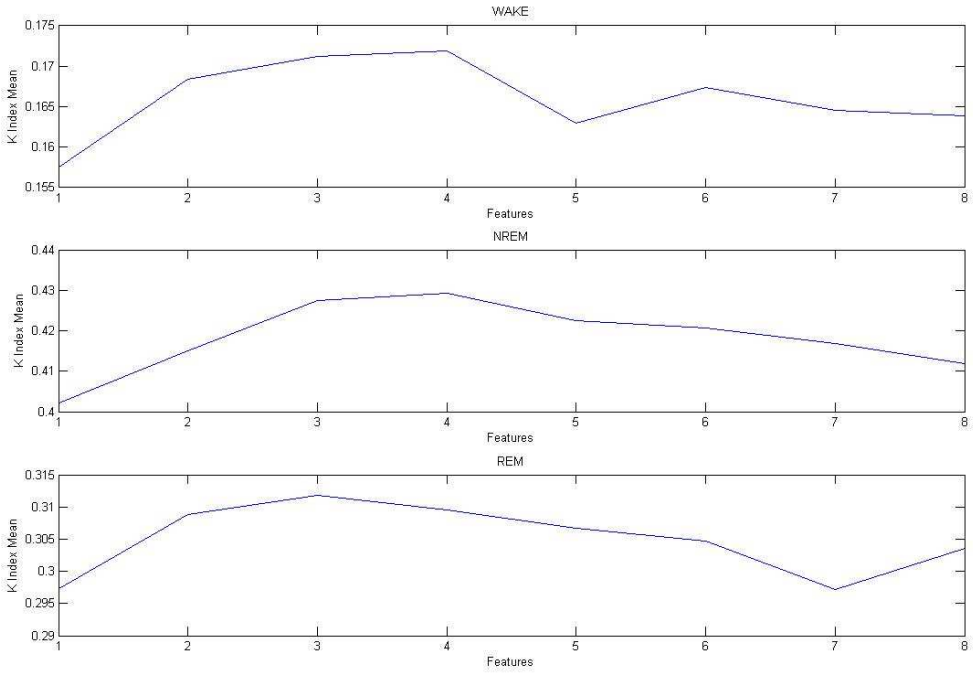


Figure 4.7: Kappa index value based on sub-optimal combination length of features extracted by RLS on RR values using a linear discriminant classifier

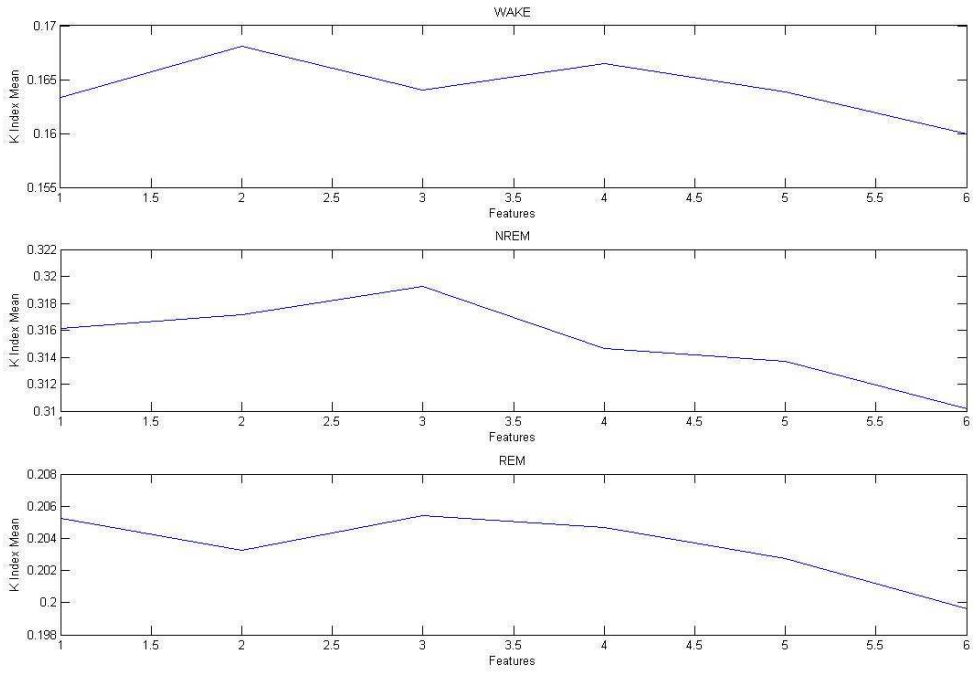


Figure 4.8: Kappa index value based on sub-optimal combination length of features extracted by RLS on respiratory values using a linear discriminant classifier

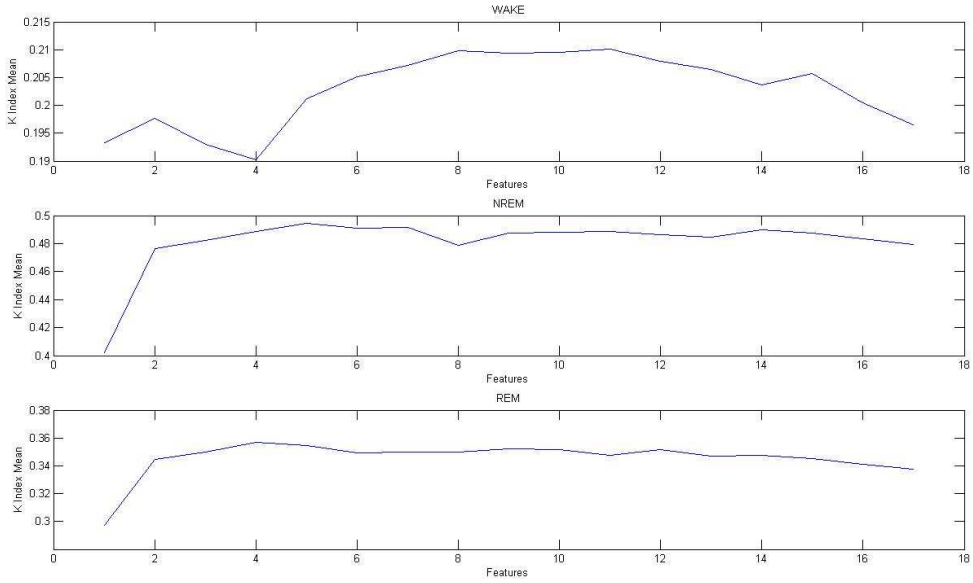


Figure 4.9: Kappa index value based on sub-optimal combination length of features extracted by RLS on RR and respiratory values using a linear discriminant classifier

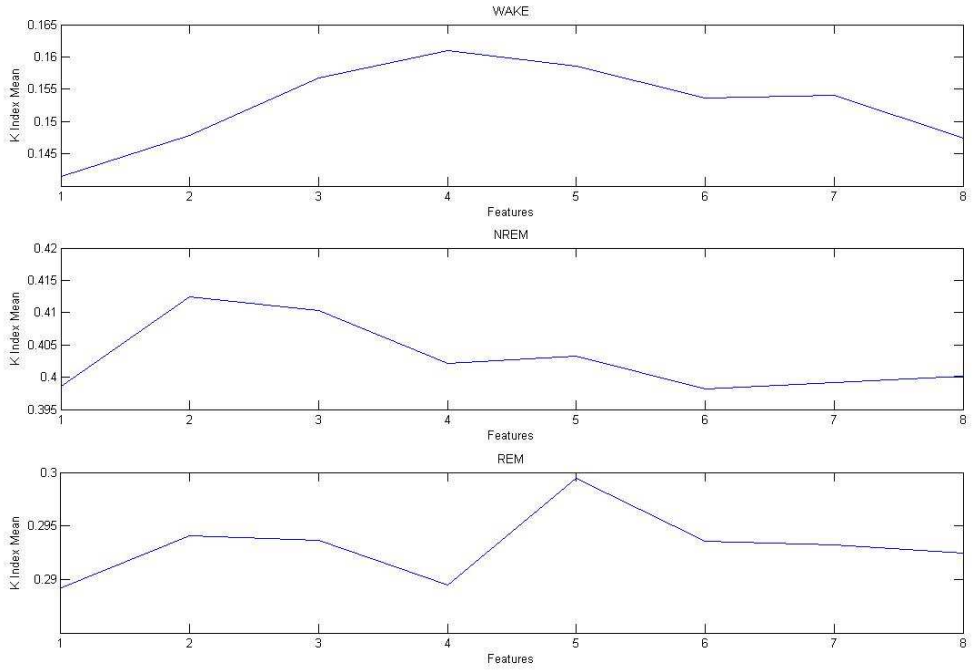


Figure 4.10: Kappa index value based on sub-optimal combination length of features extracted by RLS on RR values using a quadratic discriminant classifier

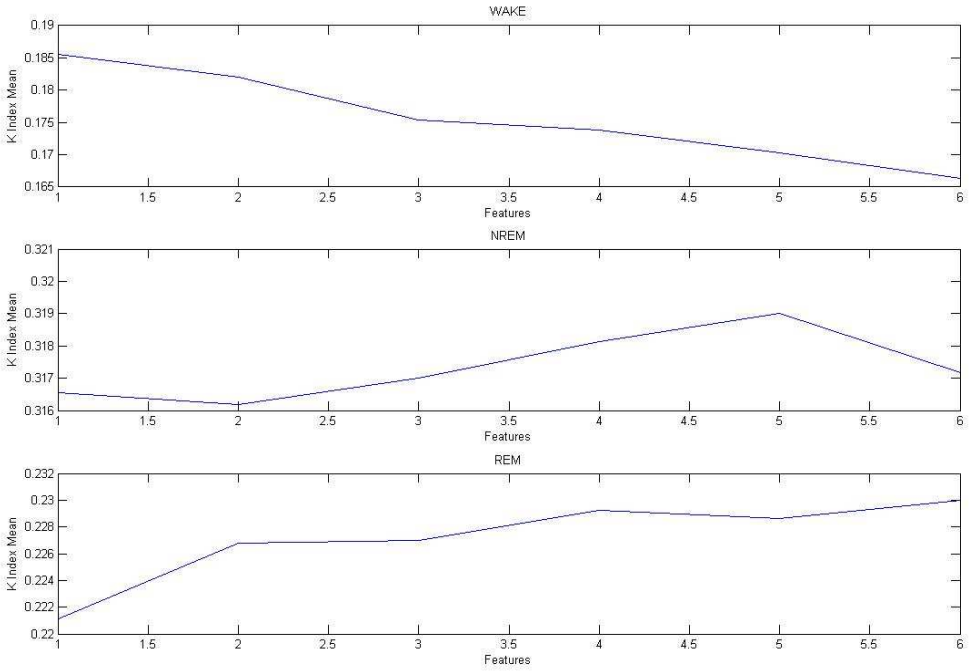


Figure 4.11: Kappa index value based on sub-optimal combination length of features extracted by RLS on respiratory values using a quadratic discriminant classifier

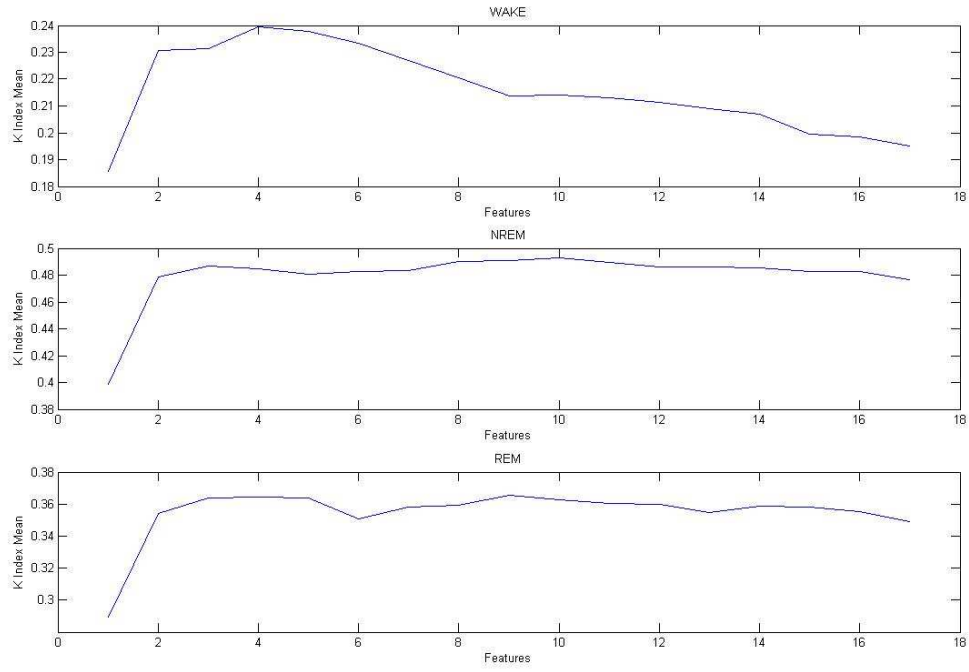


Figure 4.12: Kappa index value based on sub-optimal combination length of features extracted by RLS on RR and respiratory values using a quadratic discriminant classifier

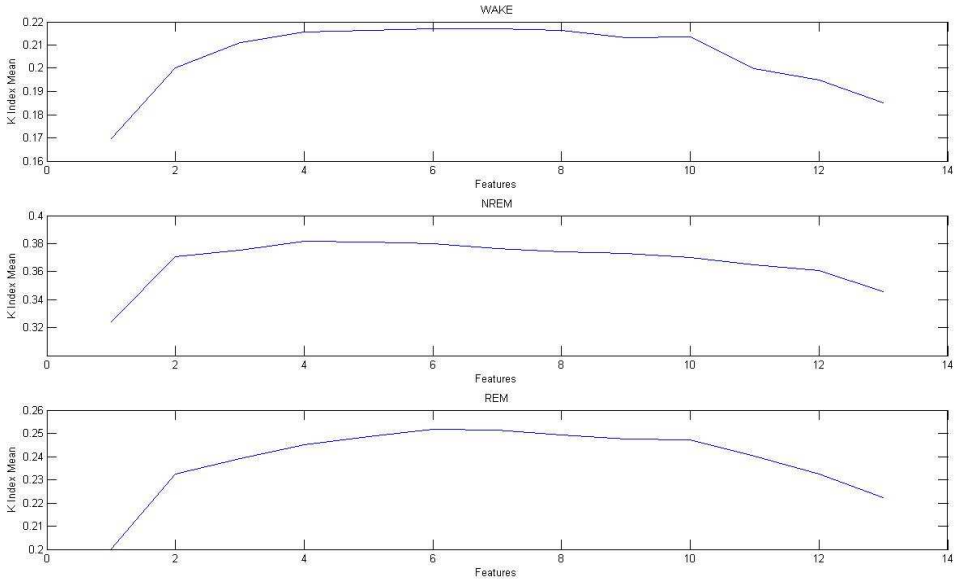


Figure 4.13: Kappa index value based on sub-optimal combination length of features extracted by wavelet on RR values using a linear discriminant classifier

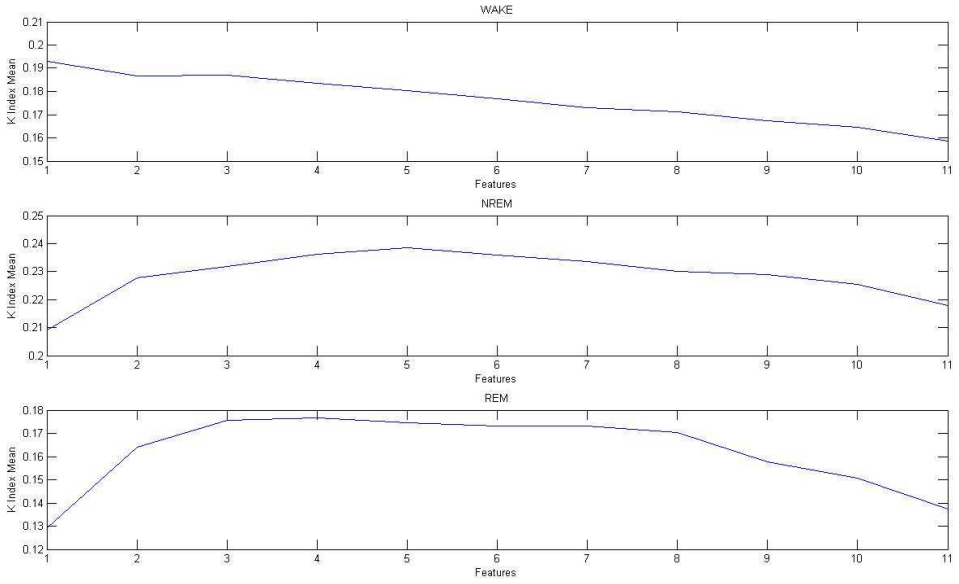


Figure 4.14: Kappa index value based on sub-optimal combination length of features extracted by wavelet on respiratory values using a linear discriminant classifier

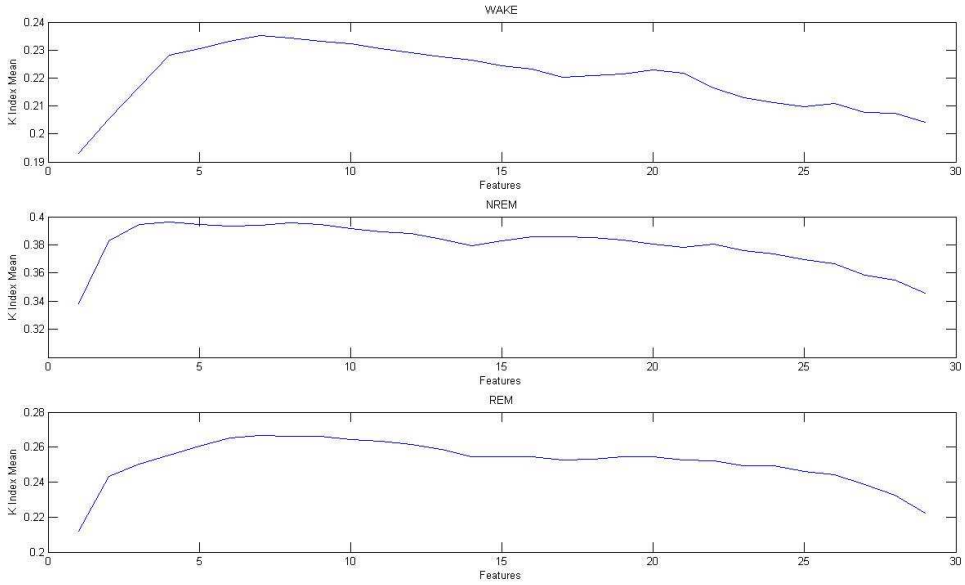


Figure 4.15: Kappa index value based on sub-optimal combination length of features extracted by wavelet on RR and respiratory values using a linear discriminant classifier

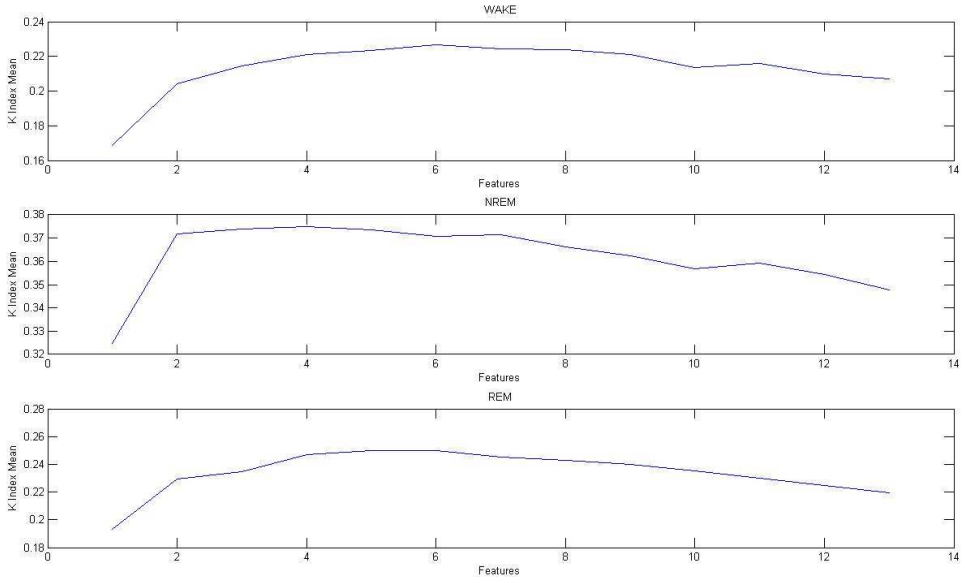


Figure 4.16: Kappa index value based on sub-optimal combination length of features extracted by wavelet on RR values using a quadratic discriminant classifier

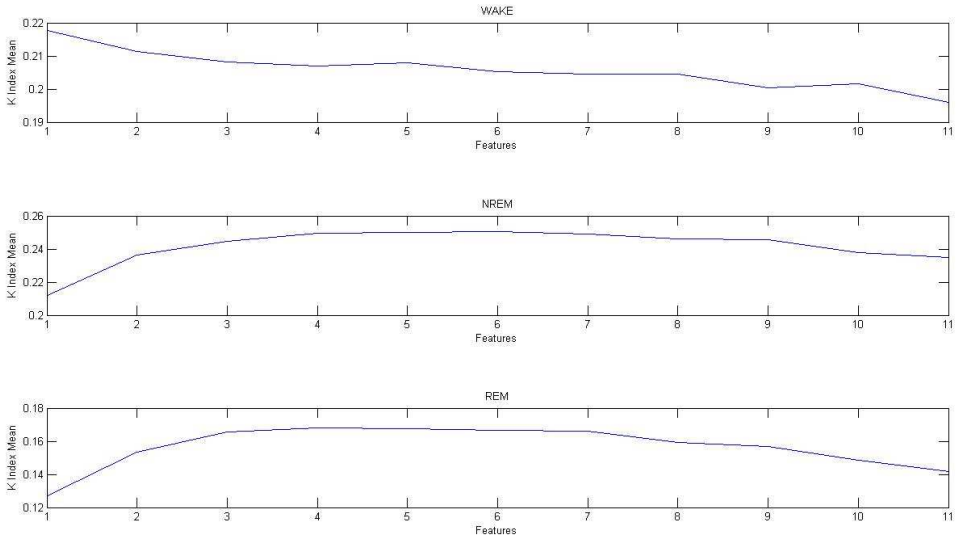


Figure 4.17: Kappa index value based on sub-optimal combination length of features extracted by wavelet on respiratory values using a quadratic discriminant classifier

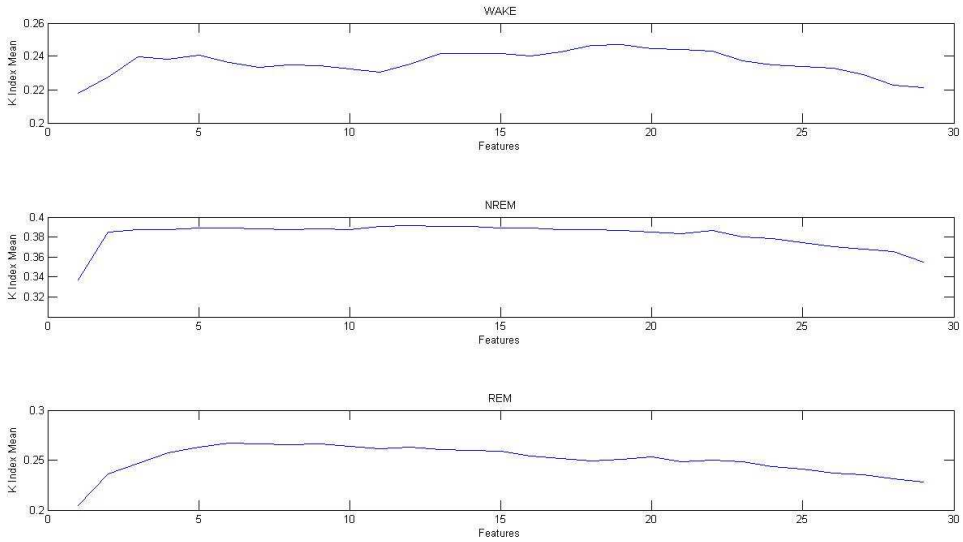


Figure 4.18: Kappa index value based on sub-optimal combination length of features extracted by wavelet on RR and respiratory values using a quadratic discriminant classifier

As one can notice, the LD and QD classifier have same performances using RR and respiratory features extracted from RLS parameters. The best kappa index is in NREM

stage. One important difference between these methods is the number of features, while QD has used 15 features derived from HRV and respiratory signals, LD has used only 8 features, which implies advantage as a reduced computational. Thus, features selected from LD are used in other classifiers. NREM has the best mean kappa index value and REM better than WAKE in each extraction method.

4.5 Feature classification and sleep stages extraction

The sleep stages were classified using a one-versus-all method with different types of classification: linear (LD) and quadratic (QD) classifier, *K*-Nearest-Neighbour (KNN) and back-propagation neural network (BPNN) method. Each indexed method uses the features selected with the sequential features selection algorithm. The three-state classification is based on two bi-state classifications: NREM versus all (REM and WAKE stages) detecting epochs with NREM stage; REM versus WAKE discriminating REM epochs and WAKE stages when epochs are not classified. As there are more NREM sleep epochs than WAKE and REM, the NREM classification has a better mean kappa index value as shown in figures from 4.7 to 4.18 (see sub-paragraph 4.4).
Figure 4.19 represents the feature selection algorithm:

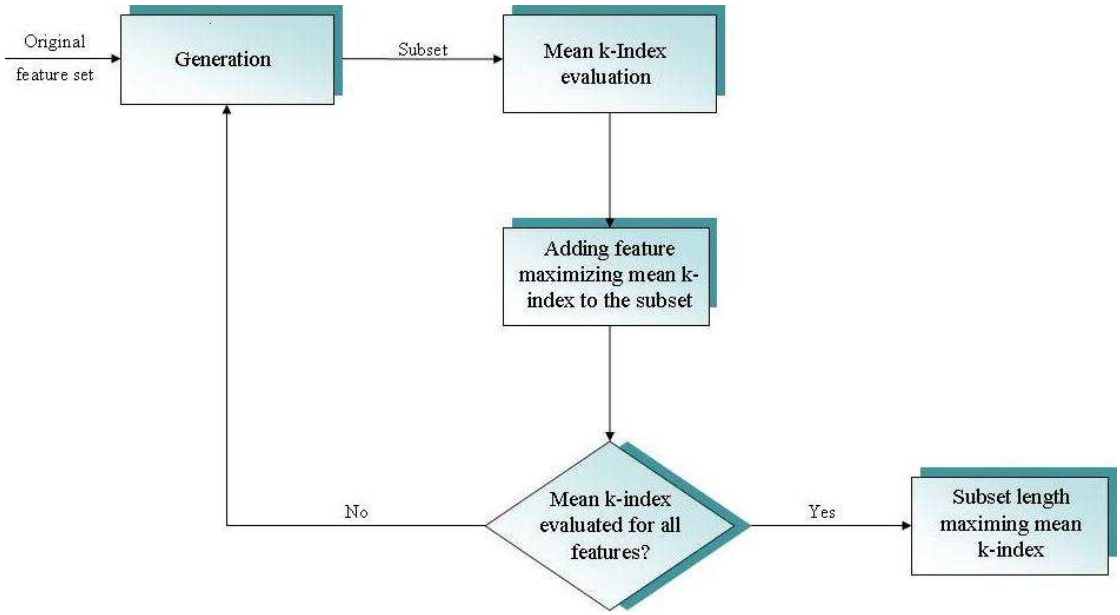


Figure 4.19: Feature selection process

The following tables show features extracted from RLS (Table 4.2) and wavelet (Table 4.3) coefficients.

Features #	Feature name	Deriving signals
1	Mean	RR
2	Standard deviation	RR
3	HF/VLF	RR
4	LF/HF	RR
5	LF/VLF	RR
6	HF/TP	RR
7	LF/TP	RR
8	VLF/TP	RR
9	HF/VLF	Respiratory
10	LF/HF	Respiratory
11	LF/VLF	Respiratory
12	HF/TP	Respiratory
13	LF/TP	Respiratory
14	VLF/TP	Respiratory
15	Cross-spectrum module of HF	RR and Respiratory
16	Cross-spectrum module of LF	RR and Respiratory
17	Cross-spectrum module of VLF	RR and Respiratory

Table 4.2: Table shows features extracted from RLS

Features #	Feature name	Deriving signals
1	Mean	RR
2	Standard deviation	RR
3	Detail 1 and 2 ratio	RR
4	Detail 1 and 3 ratio	RR
5	Detail 2 and 3 ratio	RR
6	Detail 1 and 4 ratio	RR
7	Detail 2 and 4 ratio	RR
8	Detail 3 and 4 ratio	RR

9	Detail 1 and total power ratio	RR
10	Detail 2 and total power ratio	RR
11	Detail 3 and total power ratio	RR
12	Detail 4 and total power ratio	RR
13	Detail 5 and total power ratio	RR
14	Detail 1 and 2 ratio	RR
15	Detail 1 and 3 ratio	Respiratory
16	Detail 2 and 3 ratio	Respiratory
17	Detail 1 and 4 ratio	Respiratory
18	Detail 2 and 4 ratio	Respiratory
19	Detail 3 and 4 ratio	Respiratory
20	Detail 1 and total power ratio	Respiratory
21	Detail 2 and total power ratio	Respiratory
22	Detail 3 and total power ratio	Respiratory
23	Detail 4 and total power ratio	Respiratory
24	Detail 5 and total power ratio	Respiratory
25	Cross-spectrum module of detail 1	RR and Respiratory
26	Cross-spectrum module of detail 2	RR and Respiratory
27	Cross-spectrum module of detail 3	RR and Respiratory
28	Cross-spectrum module of detail 4	RR and Respiratory
29	Cross-spectrum module of detail 5	RR and Respiratory

Table 4.3: Table shows features extracted from wavelet coefficients

4.6 Data post-processing

In some recordings the classification has shown a short REM cycle in the first sleep stages. This event is peculiar of the narcolepsy disease, but usually it lasts about five to ten minutes. However the first 5 minutes of all recordings were imposed as WAKE state since our subjects were do not suffer of narcolepsy. Hypnograms were filtered using a median filter of different degree according to extracting method and deriving signals. The order of

the filter is selected iteratively, starting search from 1 to 29 order. The best order maximizes mean kappa index for each signal was selected. The following figures show the difference between LD with optimal median filter and LD without filtering, with R&K scoring using RLS model based extraction. Kappa index is considerably improved by including also the movement activity signal after classification. In fact, detection of WAKE periods from movement activity is well accepted in clinics, where actigraphy is already a standard for insomnia diagnosis. A threshold method is used to discard the movement. Dots indicate presence of motion, thus sleep stage is forced to WAKE. When two moment events occur in a certain interval, all epochs between these are forced as WAKE stages. The research of optimal interval was performed maximizing mean kappa index value on each subject.

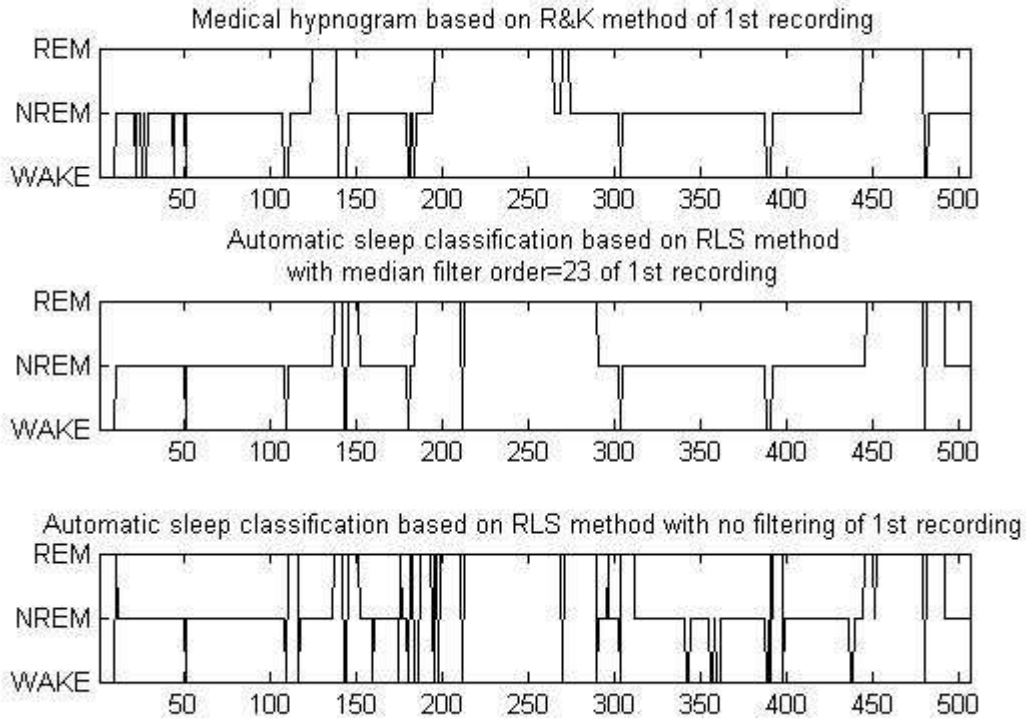


Figure 4.20: Different hypnogram estimation of 1st subject. Input signal derives from RR and respiratory signal and the features are extracted with RLS method. The upper plot shows clinical estimation of hypnogram with R&K methods. Following plots show hypnogram estimation with LD (bottom) and estimation with optimal median filter (center).

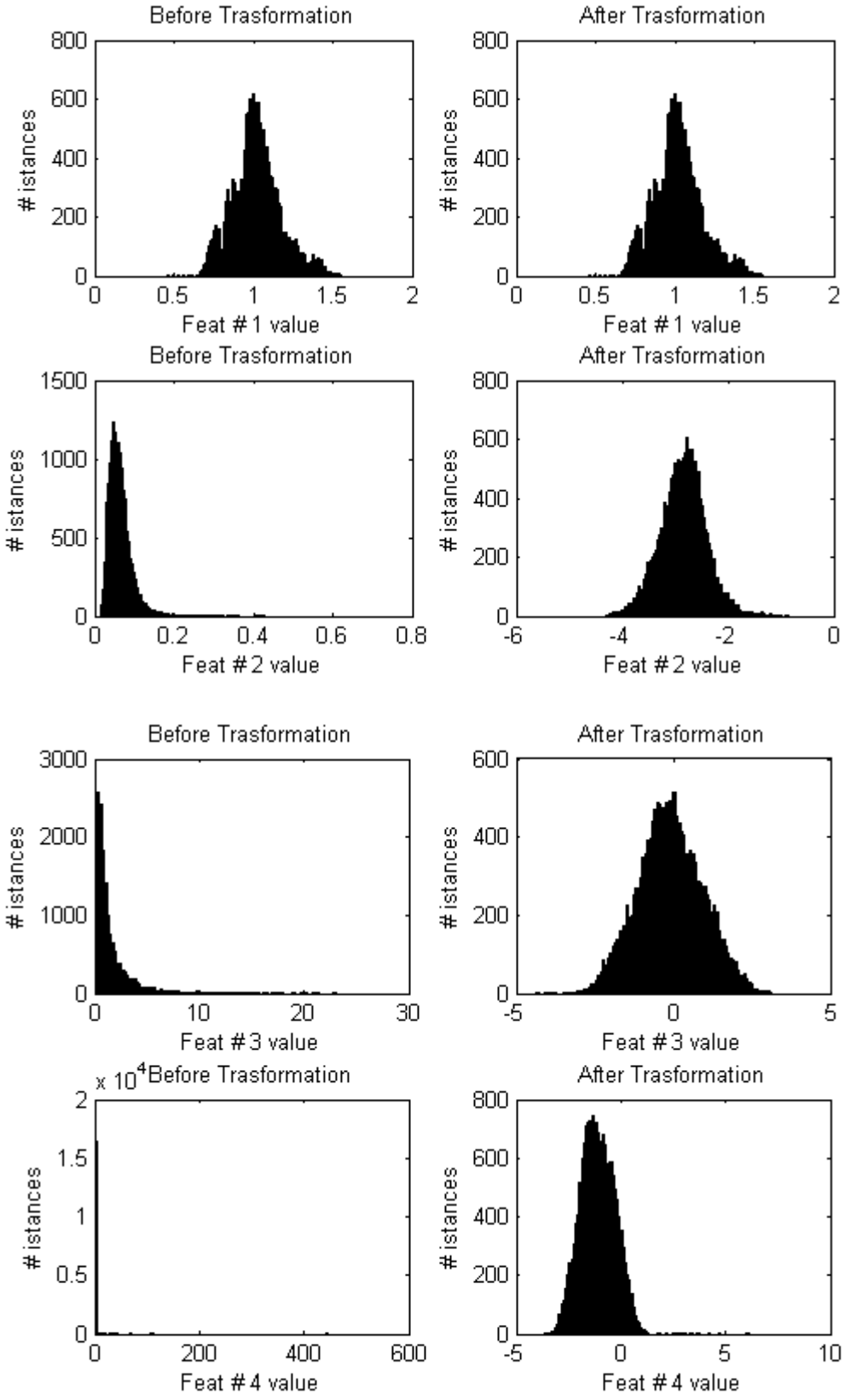
LD and QD analysis method assumes that the independent variables are normally distributed. Thus data should be transformed with most common transformation such as

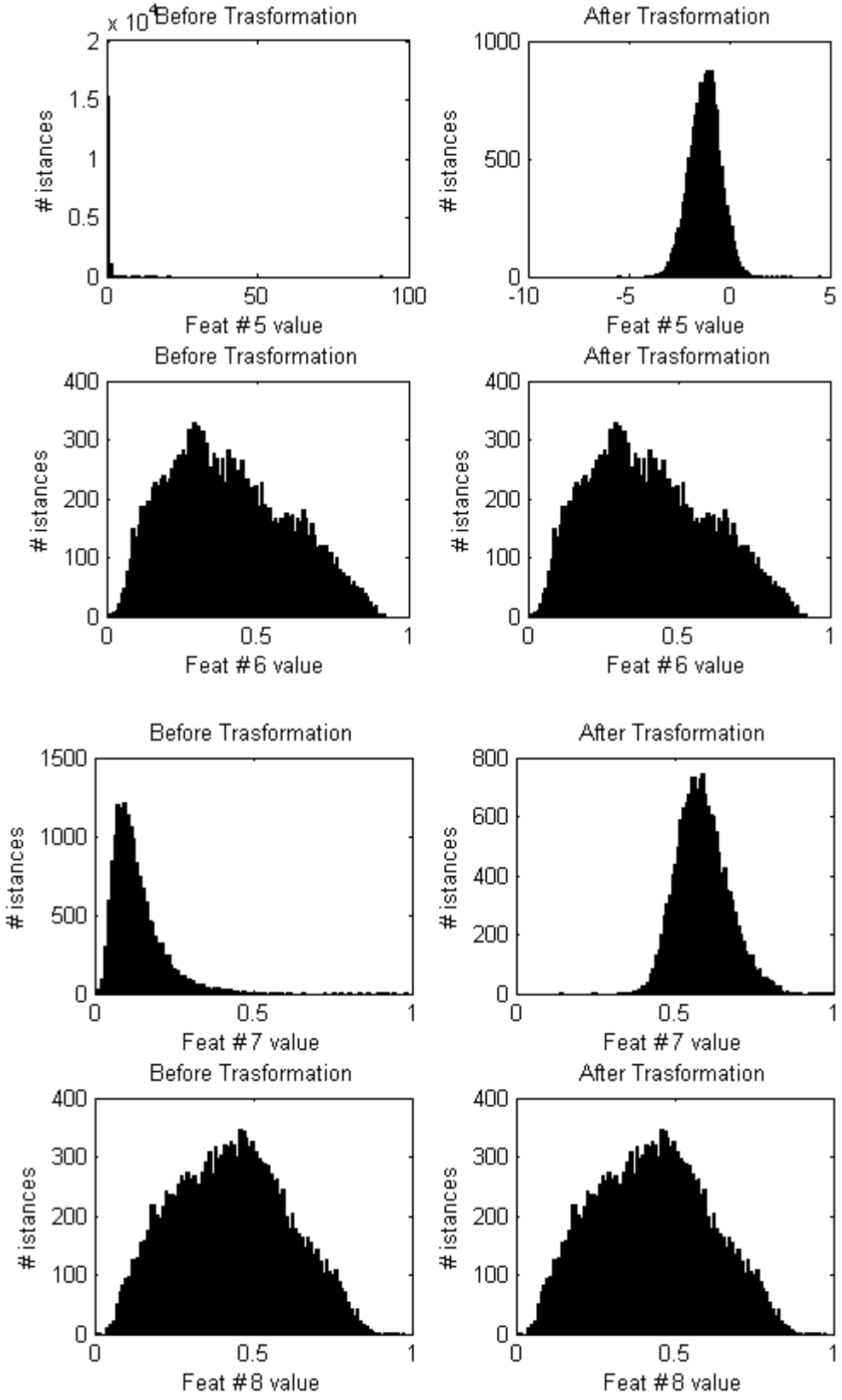
logarithmic, inversion or n^{th} root. Skewness index was used to identify distribution as normal.

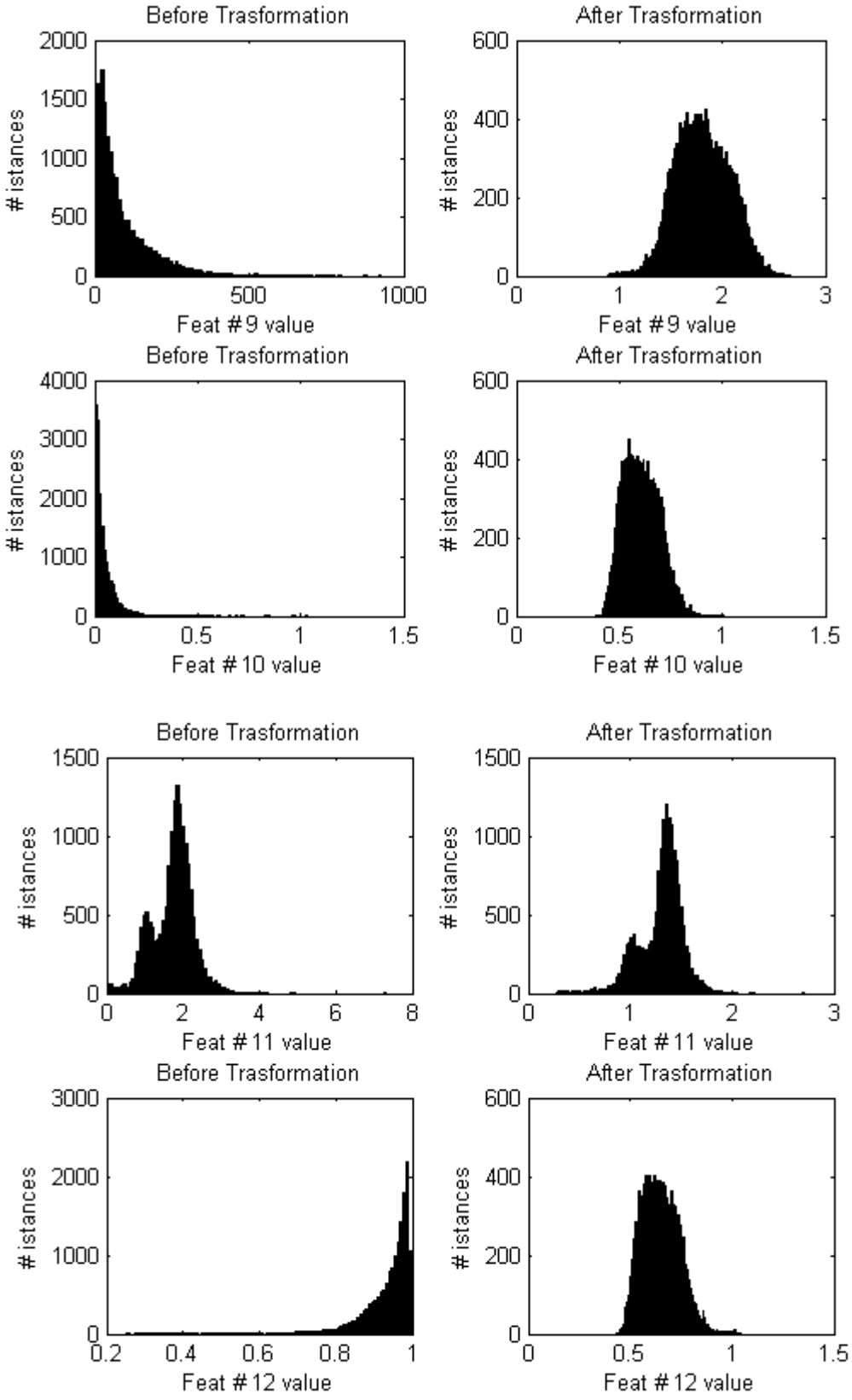
Table 4.4 and 4.5 show skewness index before and after conversion for feature extracted with RLS model (Table 4.4 and Figure 4.21) and wavelet decomposition (Table 4.5 and Figure 4.22). A normal distribution has skewness index within -1 and 1.

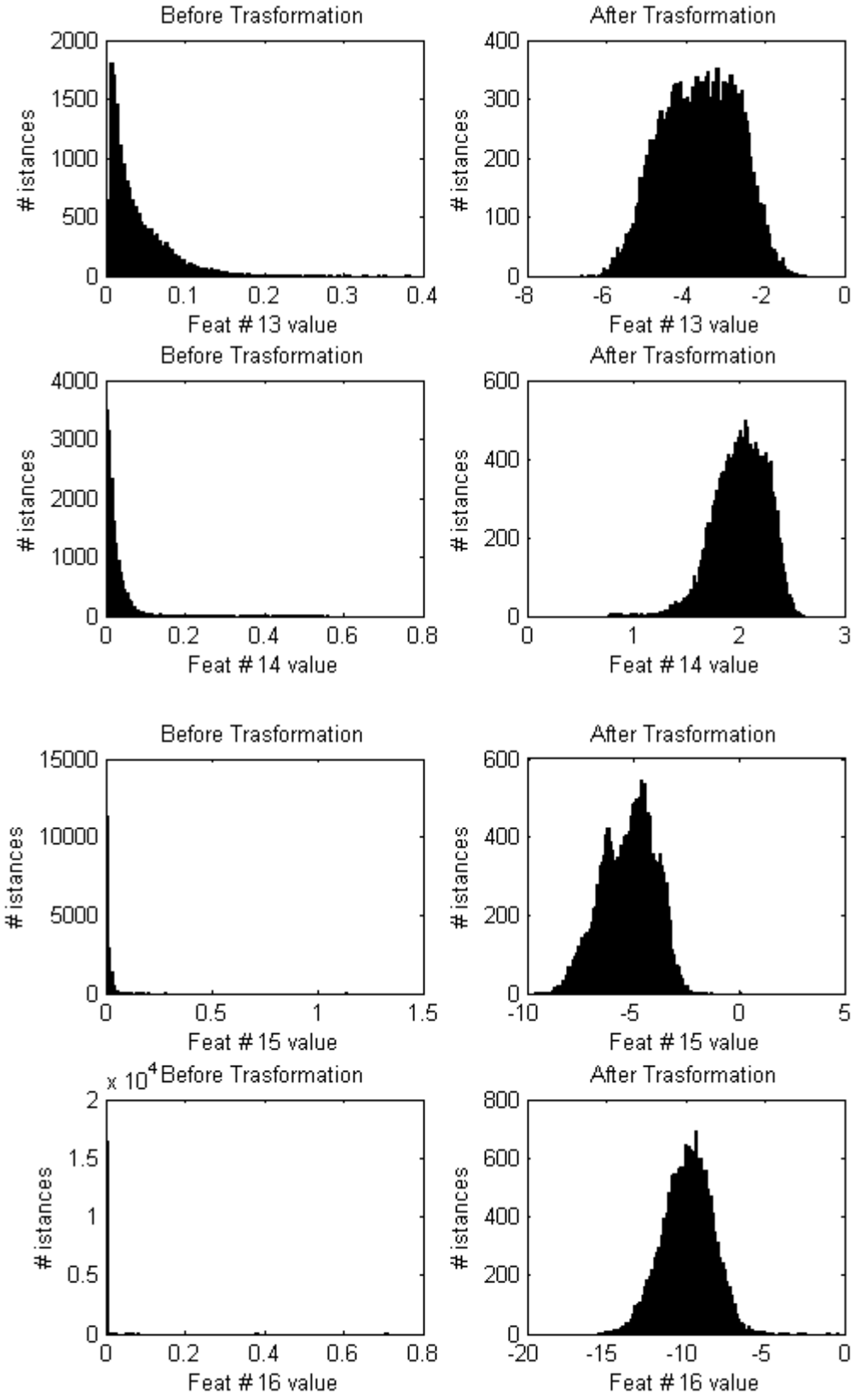
Features #	Skewness before transformation	Skewness after transformation	Transformation
1	0,382196	0,382196	Null
2	2,726889	0,096494	Logarithmic
3	3,391665	0,036104	Logarithmic
4	109,0008	0,2029	Logarithmic
5	78,38157	-0,02445	Logarithmic
6	0,344079	0,344079	Null
7	2,062712	0,41935	4^{th} root
8	0,101176	0,101176	Logarithmic
9	2,151176	0,018451	7^{th} root
10	3,968338	0,36301	7^{th} root
11	-0,1045	-0,91957	Square root
12	-3,15114	0,454705	Inversion, logarithmic and 7^{th} root
13	2,032447	-0,08484	Logarithmic
14	6,419296	-0,68712	Inversion, logarithmic and square root
15	24,69152	-0,28307	Logarithmic
16	97,24565	-0,0917	Logarithmic
17	44,18209	0,199587	Logarithmic

Table 4.4: Skewness values of features extracted with RLS method before and after transformation


Before Trasformation
After Trasformation
istances
Feat # 1 value
Feat # 2 value
Feat # 3 value
Feat # 4 value
x 10^4


Before Trasformation
x 10^4
istances
Feat #5 value
After Trasformation
istances
Feat #5 value
Before Trasformation
istances
Feat #6 value
After Trasformation
istances
Feat #6 value
Before Trasformation
istances
Feat #7 value
After Trasformation
istances
Feat #7 value
Before Trasformation
istances
Feat #8 value
After Trasformation
istances
Feat #8 value


Before Trasformation
istances
Feat #9 value
After Trasformation
istances
Feat #9 value
Before Trasformation
istances
Feat #10 value
After Trasformation
istances
Feat #10 value
Before Trasformation
istances
Feat #11 value
After Trasformation
istances
Feat #11 value
Before Trasformation
istances
Feat #12 value
After Trasformation
istances
Feat #12 value


Before Trasformation
istances
Feat # 13 value
After Trasformation
istances
Feat # 13 value
Before Trasformation
istances
Feat # 14 value
After Trasformation
istances
Feat # 14 value
Before Trasformation
istances
Feat # 15 value
After Trasformation
istances
Feat # 15 value
x 10^4 Before Trasformation
istances
Feat # 16 value
After Trasformation
istances
Feat # 16 value

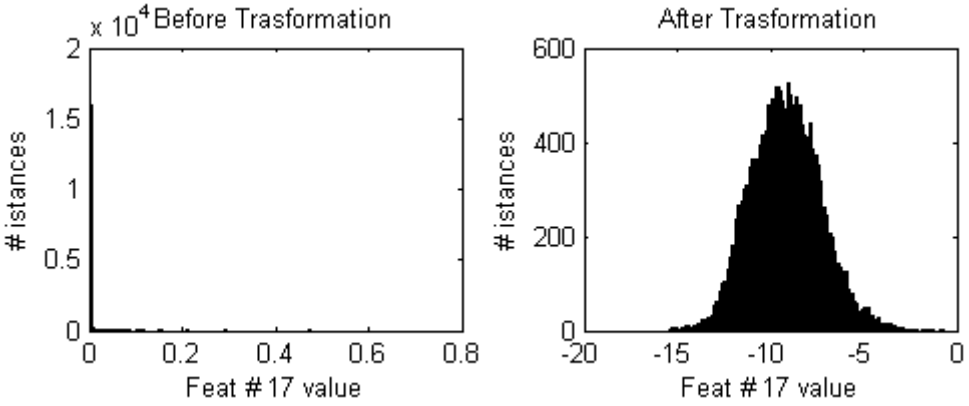
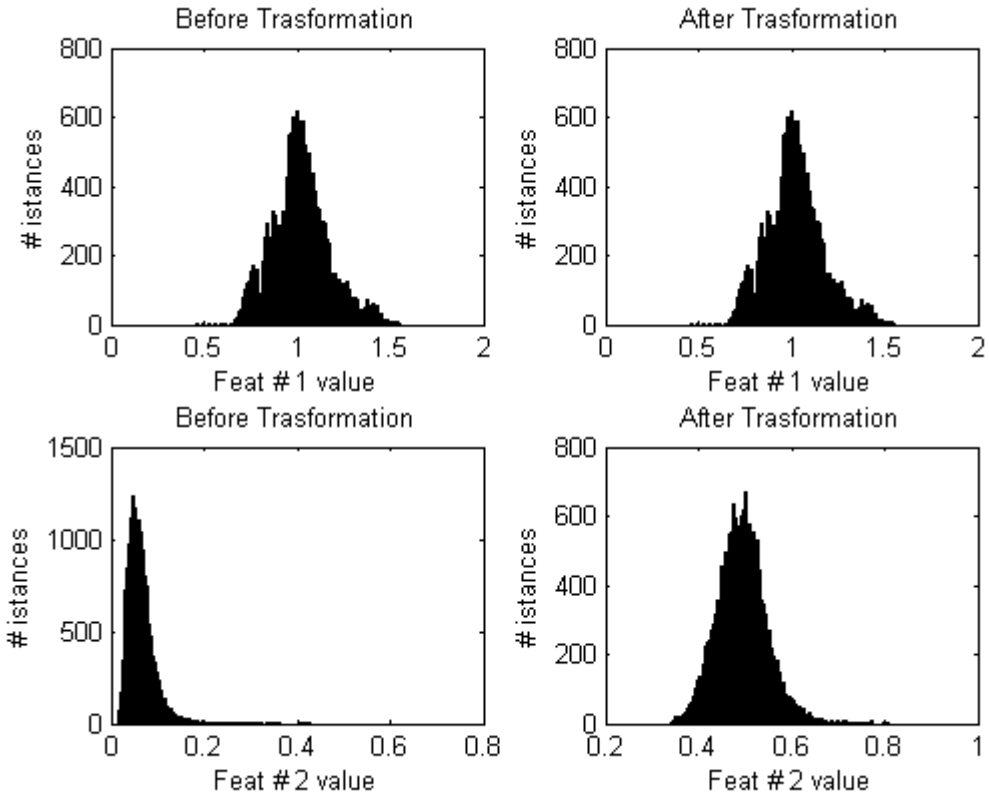


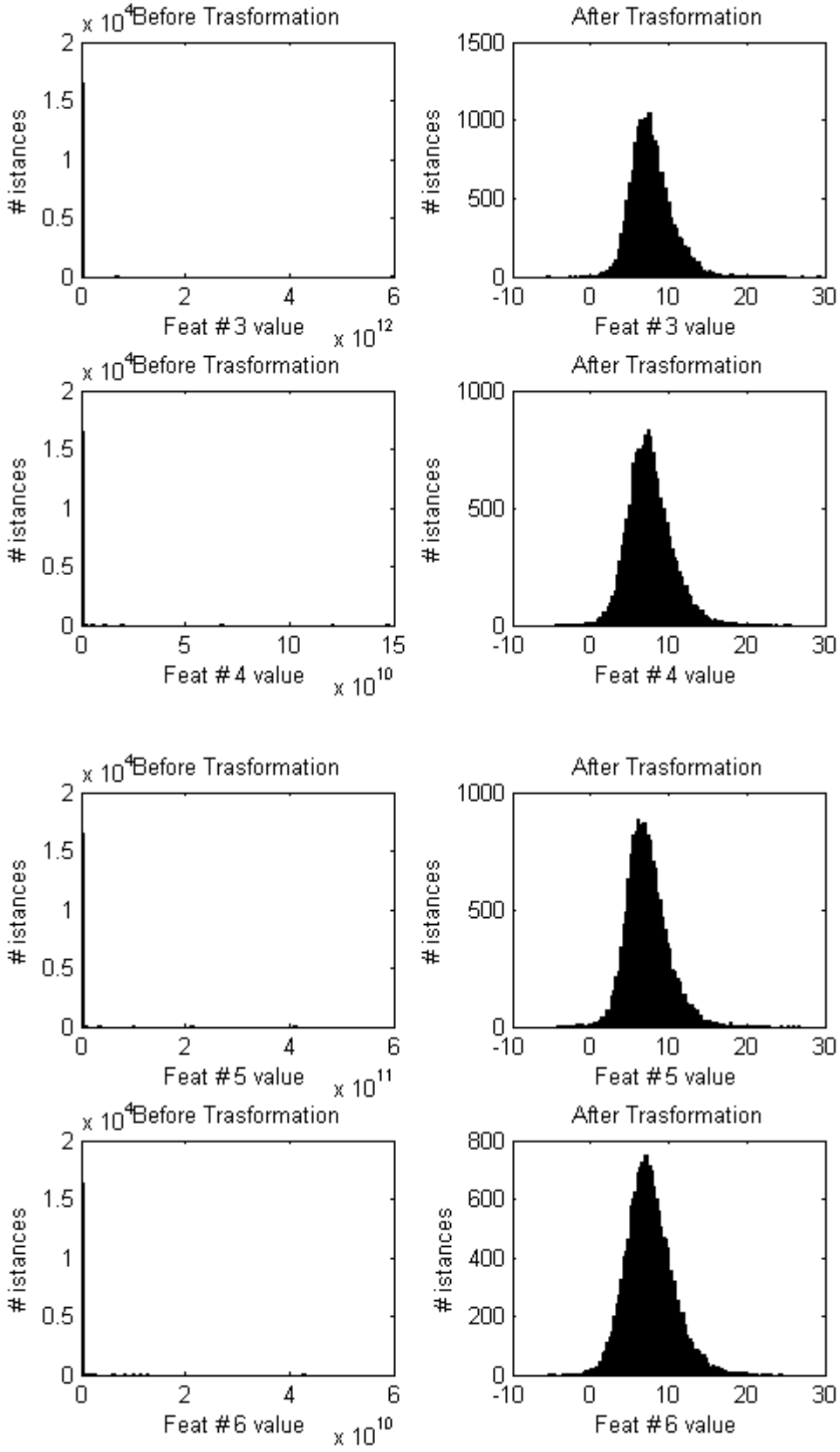
Figure 4.21: Feature distribution are shown in the figures. In the left plot the n^{th} feature is not transformed yet, while in the right plot transformation indicated in the table 4.4 is applied to n^{th} feature. Feature are extracted with RLS method.

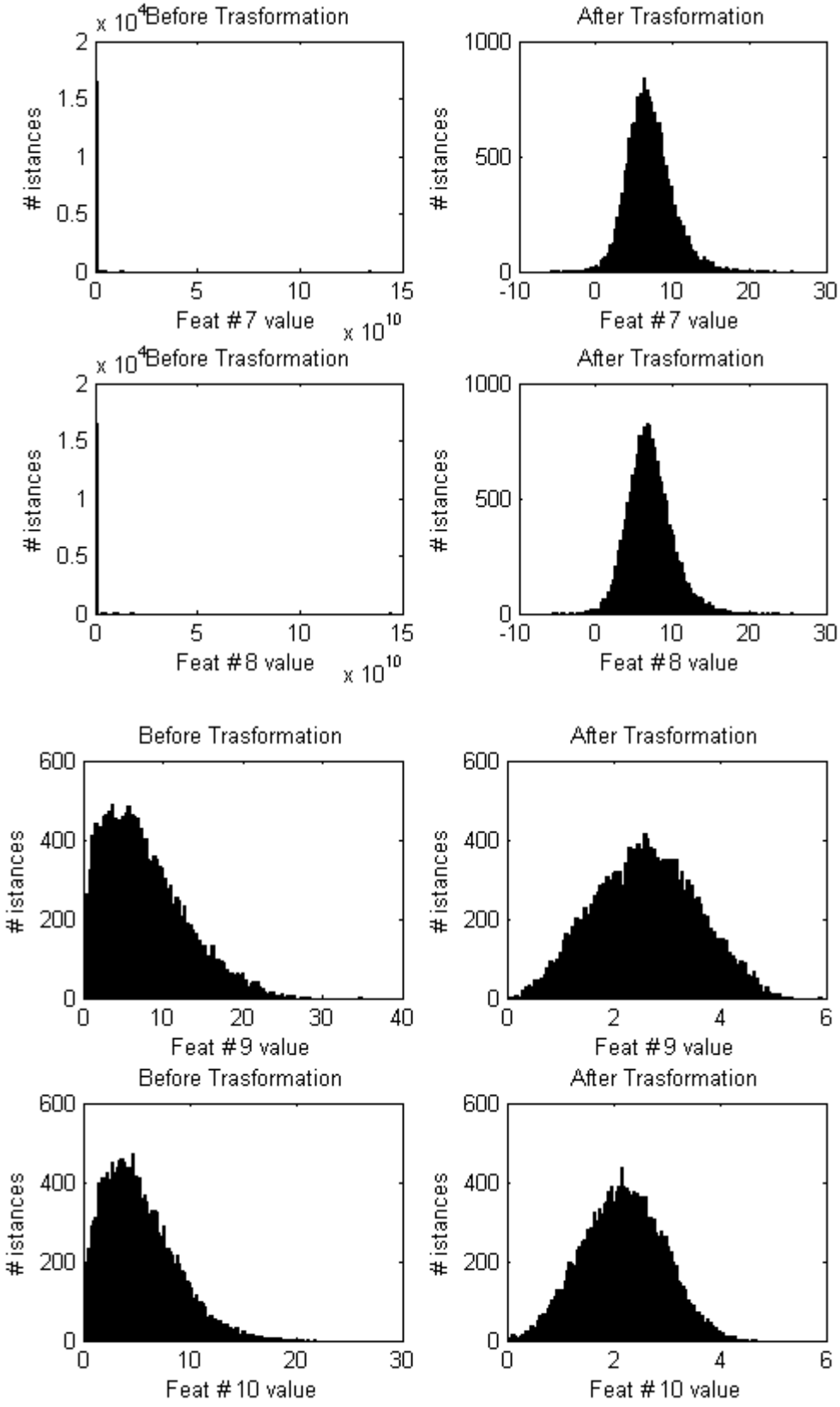
Features #	Skewness before transformation	Skewness after transformation	Transformation
1	0,382196	0,382196	Null
2	2,726889	0,562877	4^{th} root
3	125,7921	0,921831	Logarithmic
4	80,20789	0,638057	Logarithmic
5	95,7869	0,736925	Logarithmic
6	93,71481	0,535374	Logarithmic
7	126,1652	0,563087	Logarithmic
8	123,2962	0,554732	Logarithmic
9	0,821327	0,029892	Square root
10	0,865779	0,007194	Square root
11	0,927653	0,094504	Square root
12	1,091669	0,184975	Square root
13	1,620486	0,142272	3^{rd} root
14	61,19146	0,426772	Logarithmic
15	67,12076	0,272391	Logarithmic
16	105,2897	0,77544	Logarithmic

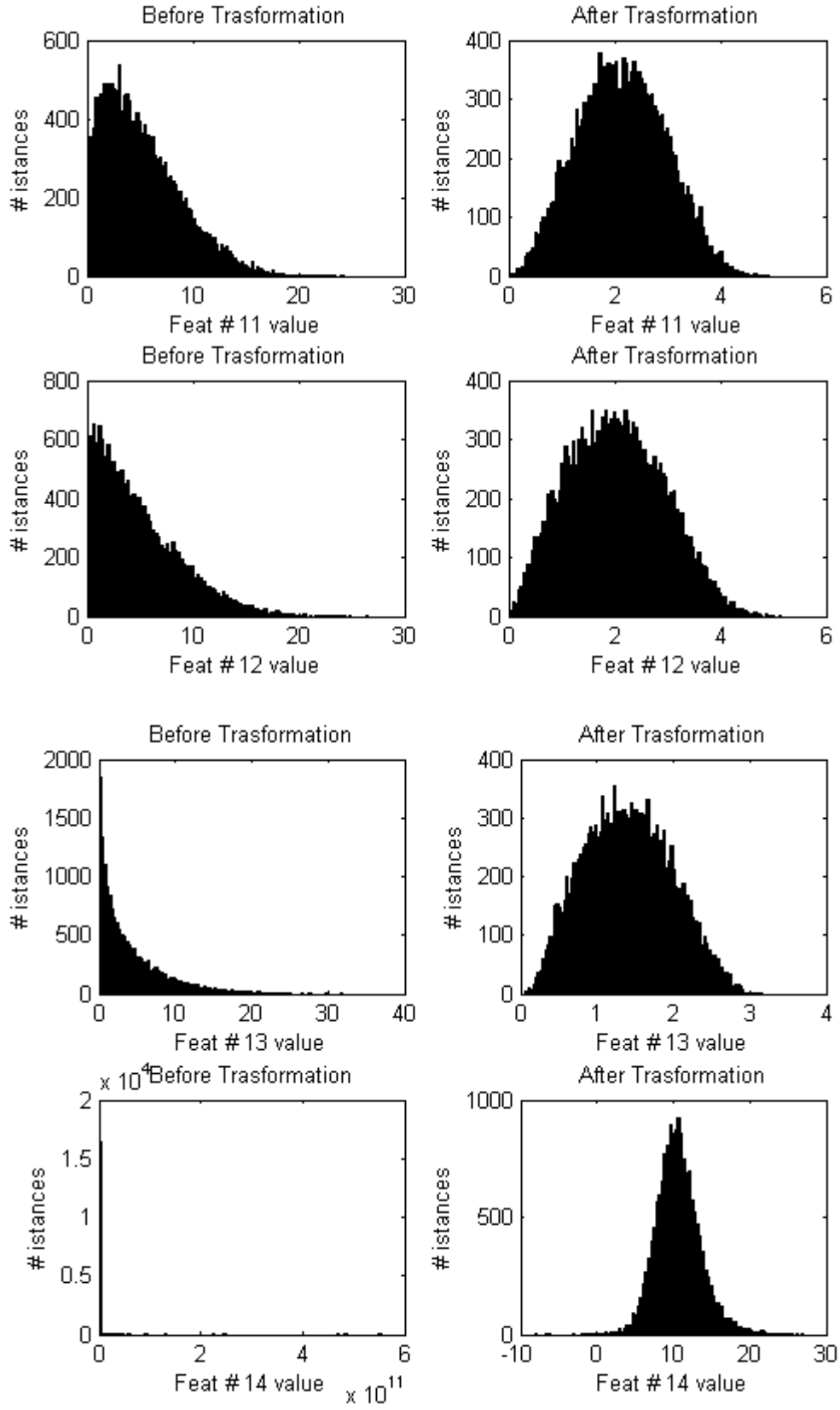
17	127,0216	0,290263	Logarithmic
18	123,4787	0,650316	Logarithmic
19	120,9149	0,588723	Logarithmic
20	-0,38916	-0,38916	Null
21	1,20698	0,182862	Square root
22	2,0282	0,248606	3^{rd} root
23	3,076858	-0,69456	Logarithmic
24	4,706542	-0,59544	Logarithmic
25	0,923615	0,079794	Square root
26	2,179225	0,024851	4^{th} root
27	3,493947	-0,71423	Logarithmic
28	4,903038	-0,57876	Logarithmic
29	8,589227	-0,51141	Logarithmic

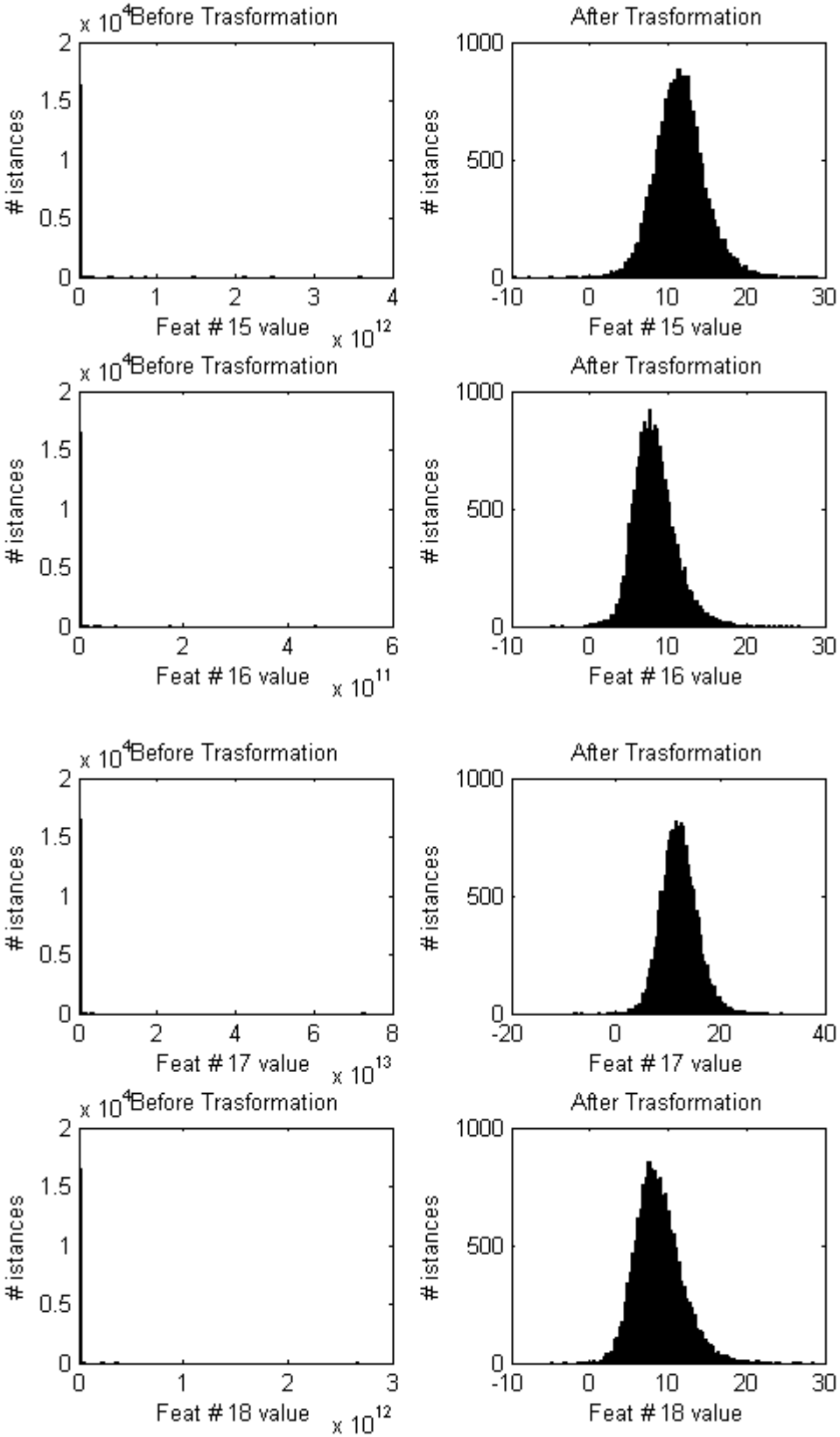
Table 4.5: Skewness values of features extracted with wavelet decomposition before and after transformation

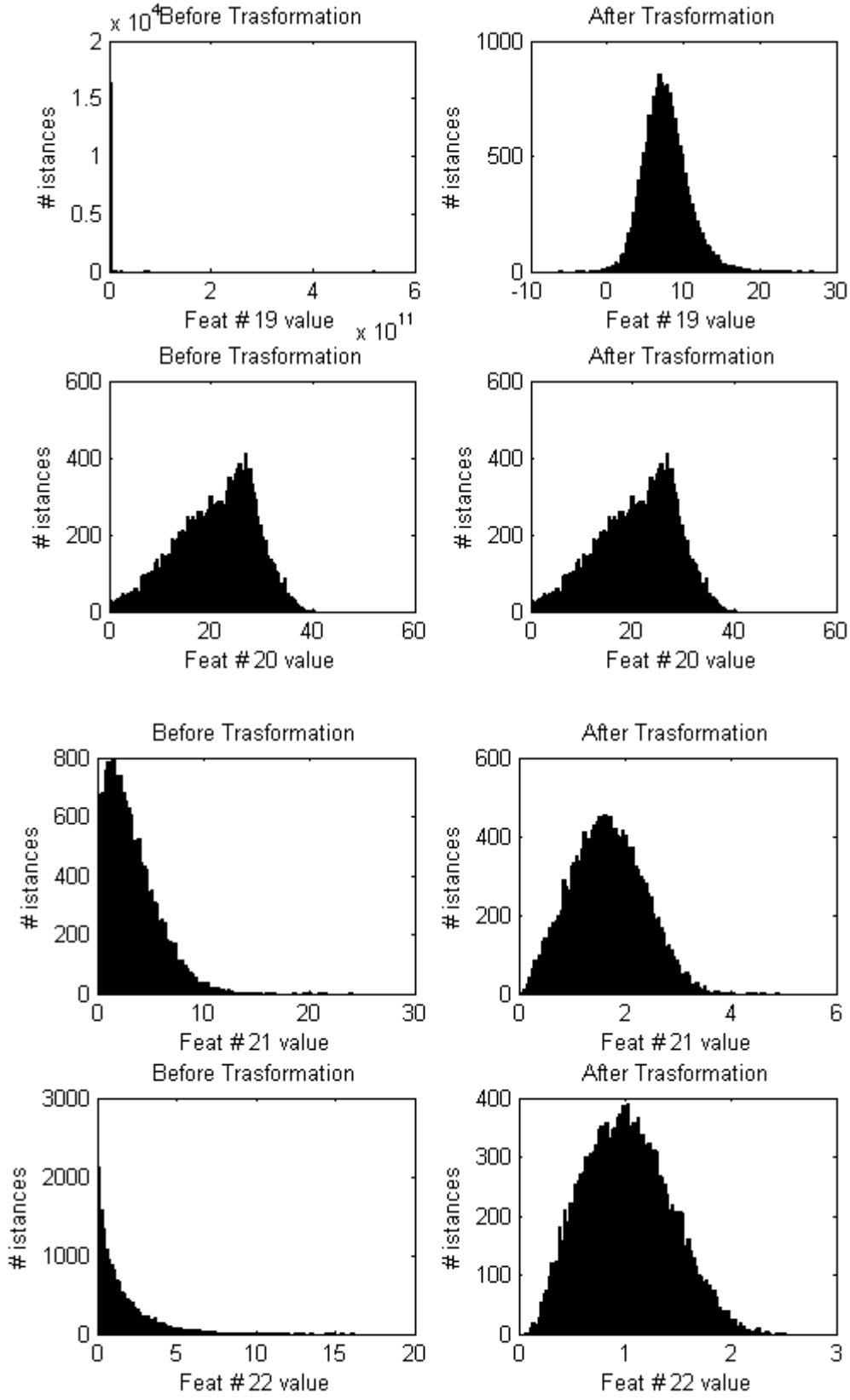


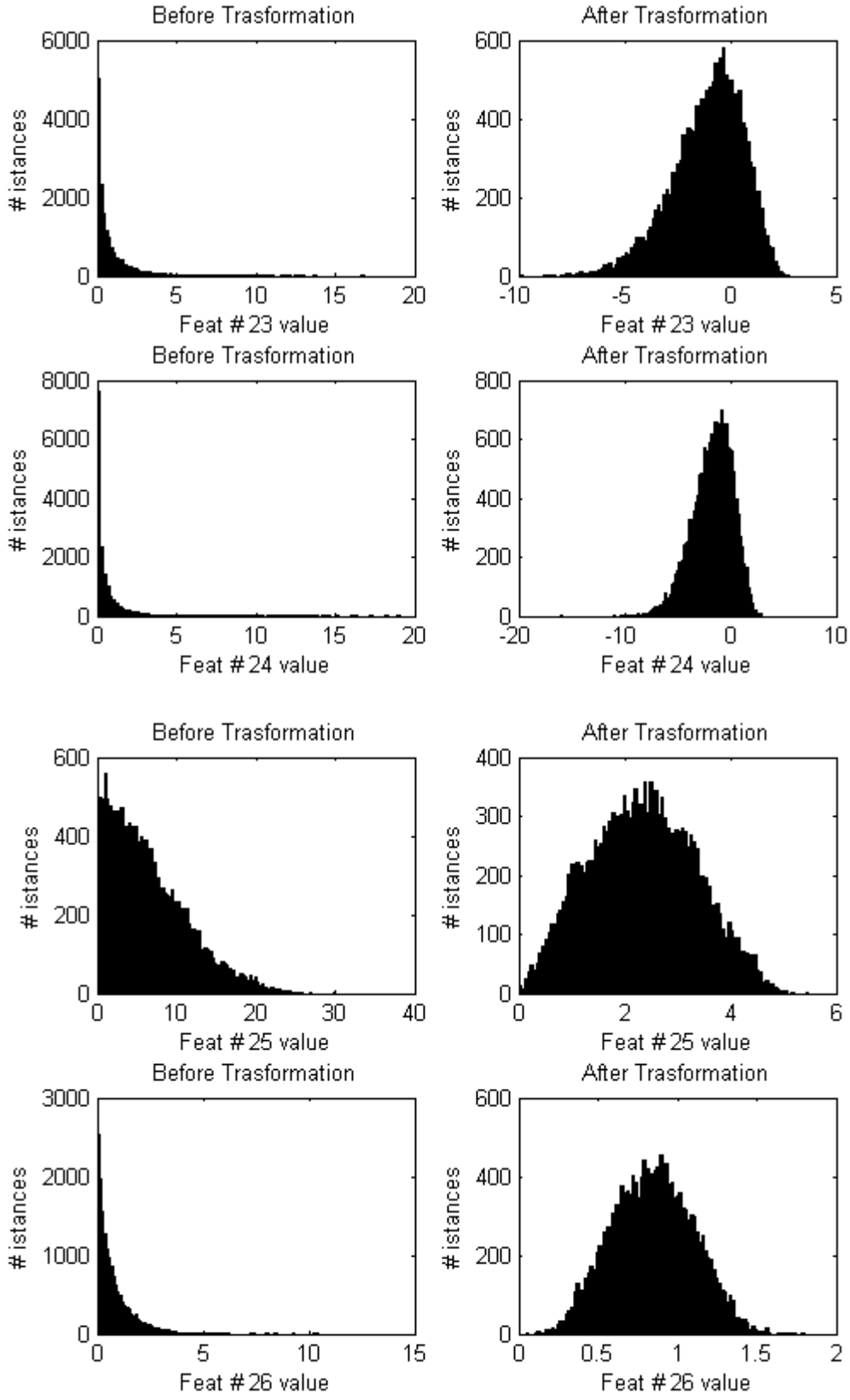

x 10^4 Before Trasformation
istances
Feat #3 value
x 10^12
After Trasformation
istances
Feat #3 value
x 10^4 Before Trasformation
istances
Feat #4 value
x 10^10
After Trasformation
istances
Feat #4 value
x 10^4 Before Trasformation
istances
Feat #5 value
x 10^11
After Trasformation
istances
Feat #5 value
x 10^4 Before Trasformation
istances
Feat #6 value
x 10^10
After Trasformation
istances
Feat #6 value


x 10^4 Before Trasformation
istances
Feat #7 value
x 10^10
After Trasformation
istances
Feat #7 value
x 10^4 Before Trasformation
istances
Feat #8 value
x 10^10
After Trasformation
istances
Feat #8 value
Before Trasformation
istances
Feat #9 value
After Trasformation
istances
Feat #9 value
Before Trasformation
istances
Feat #10 value
After Trasformation
istances
Feat #10 value


Before Trasformation
istances
Feat # 11 value
After Trasformation
istances
Feat # 11 value
Before Trasformation
istances
Feat # 12 value
After Trasformation
istances
Feat # 12 value
Before Trasformation
istances
Feat # 13 value
After Trasformation
istances
Feat # 13 value
x 10^4 Before Trasformation
istances
Feat # 14 value
x 10^11
After Trasformation
istances
Feat # 14 value


x 10^4 Before Trasformation
istances
Feat # 15 value
x 10^12
After Trasformation
istances
Feat # 15 value
x 10^4 Before Trasformation
istances
Feat # 16 value
x 10^11
After Trasformation
istances
Feat # 16 value
x 10^4 Before Trasformation
istances
Feat # 17 value
x 10^13
After Trasformation
istances
Feat # 17 value
x 10^4 Before Trasformation
istances
Feat # 18 value
x 10^12
After Trasformation
istances
Feat # 18 value


x 10^4 Before Trasformation
istances
Feat # 19 value
After Trasformation
istances
Feat # 19 value
x 10^11 Before Trasformation
istances
Feat # 20 value
After Trasformation
istances
Feat # 20 value
Before Trasformation
istances
Feat # 21 value
After Trasformation
istances
Feat # 21 value
Before Trasformation
istances
Feat # 22 value
After Trasformation
istances
Feat # 22 value


Before Trasformation
After Trasformation
istances
Feat # 23 value
Feat # 24 value
Feat # 25 value
Feat # 26 value

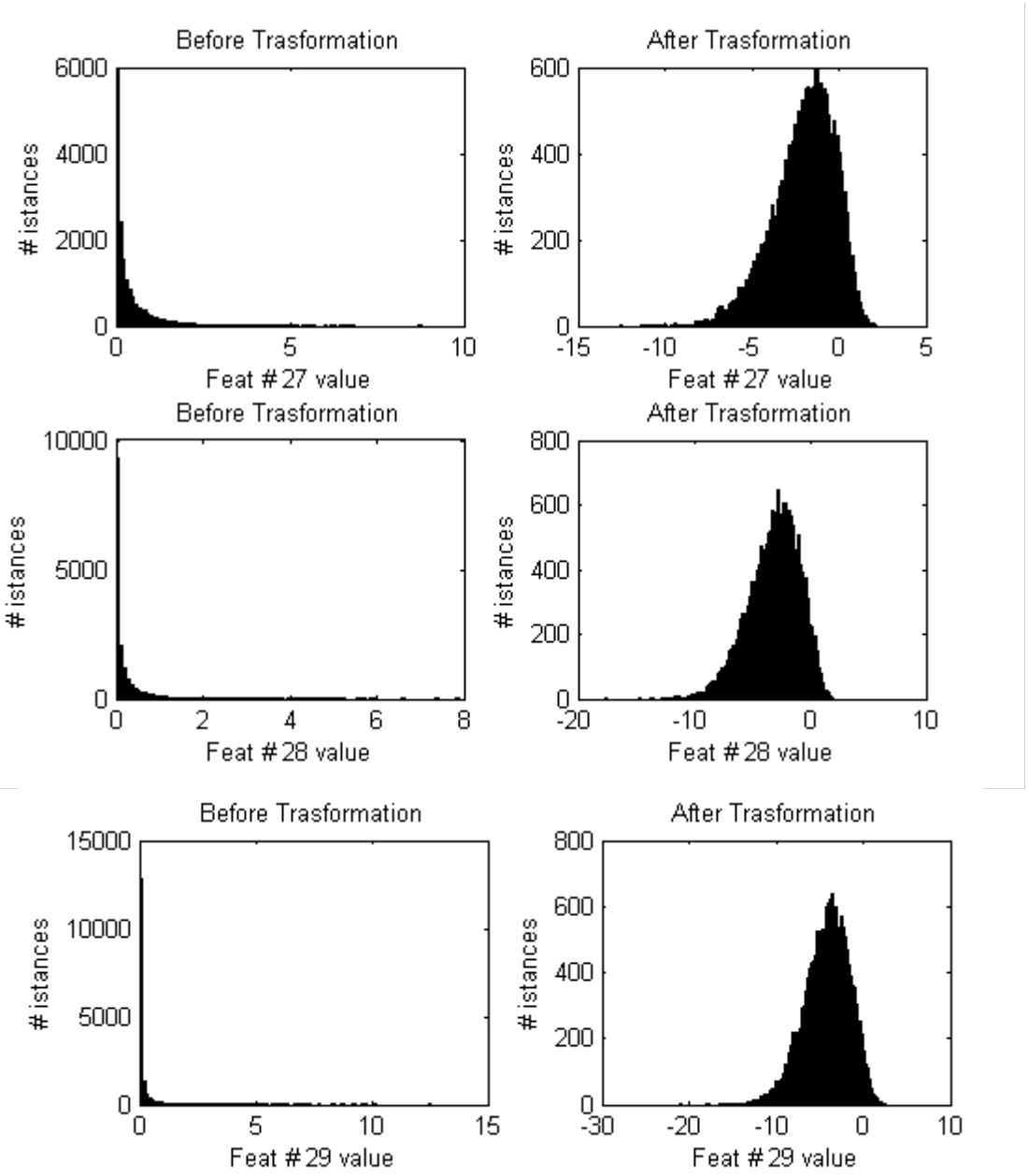


Figure 4.22: Feature distribution are shown in the figures. In the left plot the feature n^{th} is not transformed yet, while in the right plot transformation indicated in the table 4.5 is applied to feature n^{th}. Feature are extracted with wavelet coefficients.

Chapter 5

Results

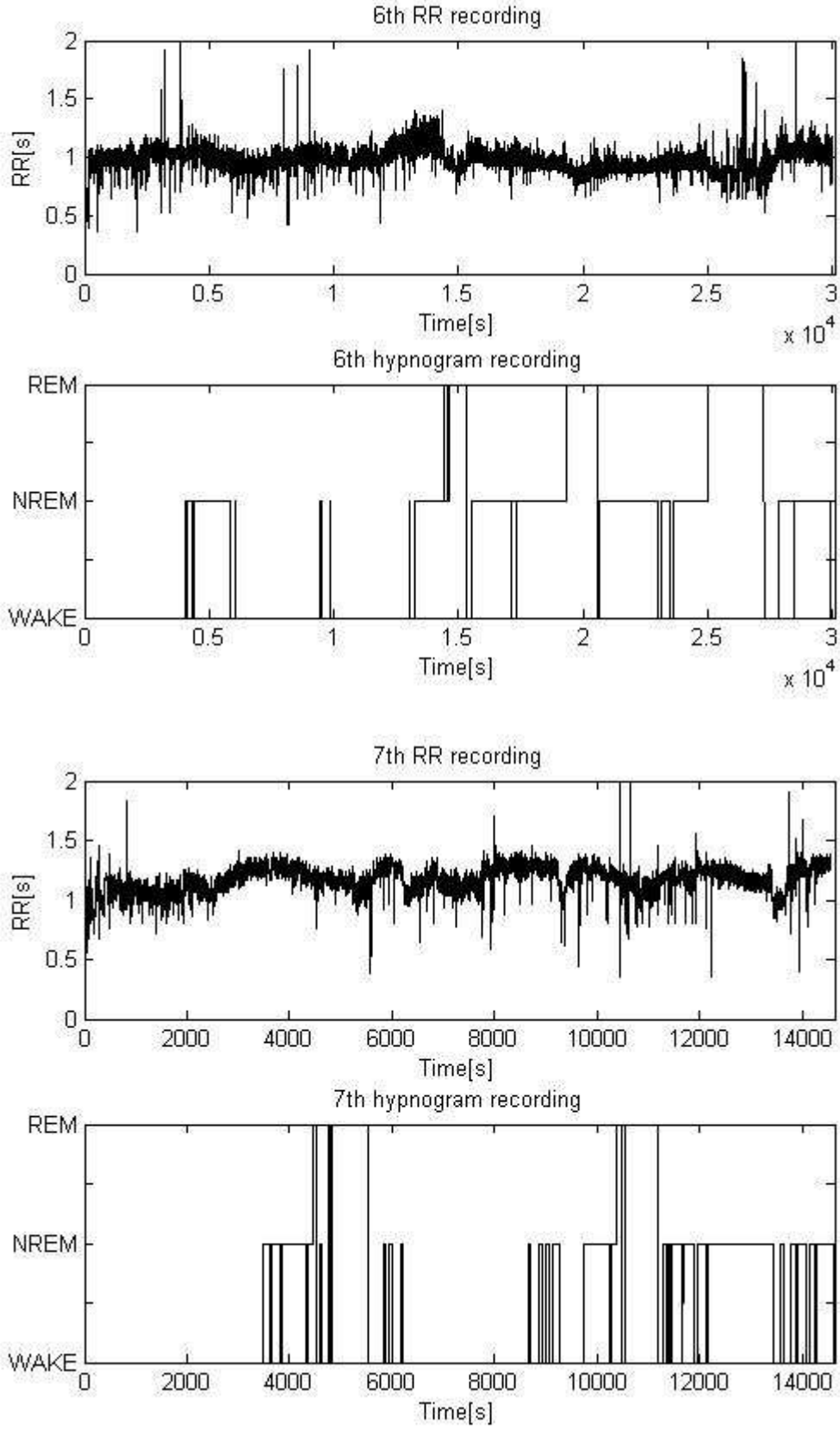
"There are two possible outcomes: if the result confirms the hypothesis, then you've made a measurement. If the result is contrary to the hypothesis, then you've made a discovery."

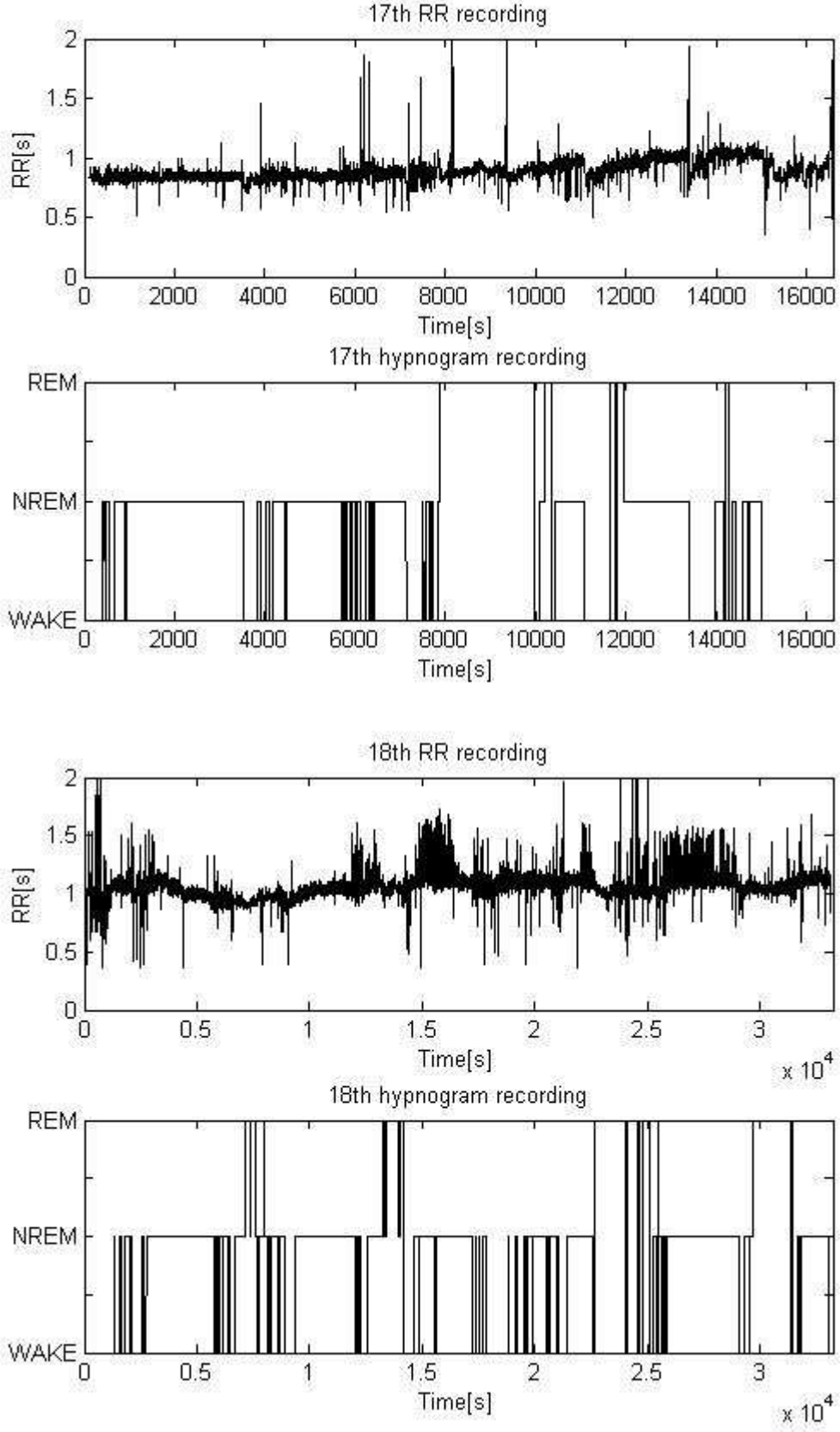
Enrico Fermi

In this chapter, results of previously described methods (Linear Discriminant, Quadratic Discriminant, *K*-Nearest-Neighbor and Neural Network) are listed. We use three different type of signals for each method: RR; respiratory; RR and respiratory together. We defined that the presence of movement activity classifies stages as WAKE. The analysis was computed by classifying the final sleep profile: with movement activity, without movement activity or with movement activity and optimal median filter. The best computing method is chosen by a comparison between all these different cases (different combination of signal nature – HRV, Respiration and movement - and their combination) for each classifier.

An additional procedure for wake stage was done as follows: when two movement events occur in a certain interval, all epochs between these are forced as WAKE stages. Median filter order and movement-wake interval were chosen iteratively by maximizing mean Kappa index value on every subject afterwards, the mean median filter order and movement-wake interval were used for all the subjects.

Recording 6, 7, 17, 18 and 21 are not used because they presented a reduced heart rate variability and altered mean heart frequency as shown in the figure 5.1.


6th RR recording
RR[s]
Time[s]
x 10^4
6th hypnogram recording
REM
NREM
WAKE
Time[s]
x 10^4
7th RR recording
RR[s]
Time[s]
7th hypnogram recording
REM
NREM
WAKE
Time[s]


17th RR recording
RR[s]
Time[s]
17th hypnogram recording
REM
NREM
WAKE
Time[s]
18th RR recording
RR[s]
Time[s]
x 10^4
18th hypnogram recording
REM
NREM
WAKE
Time[s]
x 10^4

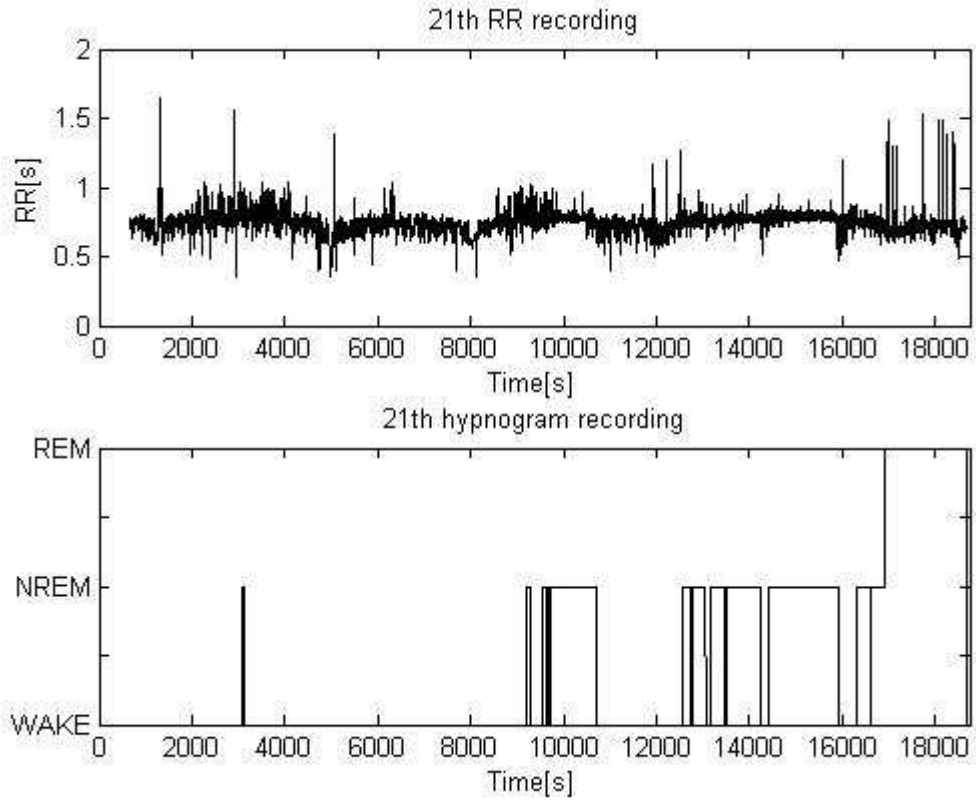


Figure 5.1: Recording 6 and 7 have altered mean heart frequency, while recording 17, 18 and 21 presents also reduced heart variability.

The algorithms have been developed using MATLAB 7.4. The experiments were run on a 3.00GHz Intel® Pentium IV processor with 2048MB of RAM under Windows XP. The recognition procedure is performed off-line on data stored on the computer's hard disk. The performance of the automated recognition system is determined by measuring accuracy and Cohen's Kappa Coefficient (k). k is a measure of inter-rate agreement, where the two rates are the expert sleep technician (who scored the polysomnography recordings) and the automated sleep staging system (it can be vary from k=1 for perfect agreement to k=0 for a performance no better than chance).

5.1 Linear Discriminant

Linear Discriminant was used to select features (see Chapter 4.4) and to classify sleep stages. In this chapter classification with LD is described. It was selected the optimal

WAKE interval and median filter degree selecting iteratively starting from 1 to 10 and from 1 to 29, respectively, as maximizing mean kappa index value, as shown in figures from 5.2 to 5.7, which showing mean kappa index.

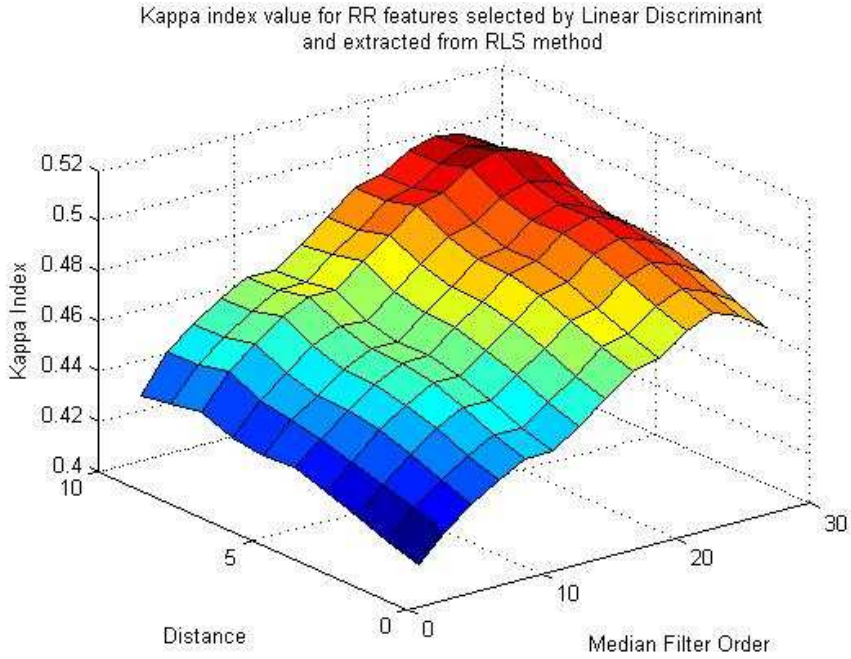


Figure 5.2: Median filter degree and optimal distances based on mean kappa index for each subject computing with LD and using RR features extracted with RLS method. Median filter order=25 and distance=8 maximize mean kappa index value.

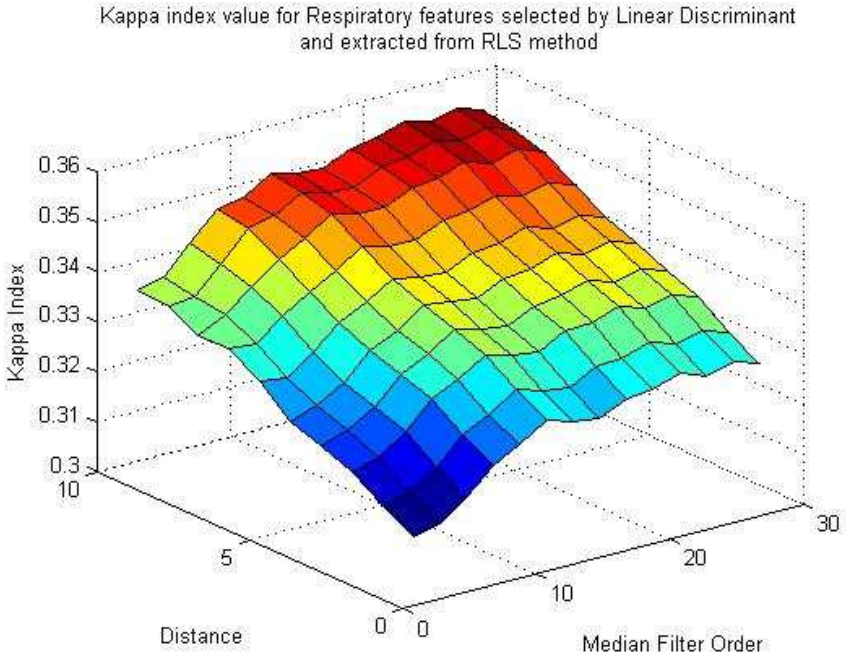


Figure 5.3: Median filter degree and optimal distances based on mean kappa index for each subject computing with LD and using respiratory features extracted with RLS method Median filter order=25 and distance=8 maximize mean kappa index value.

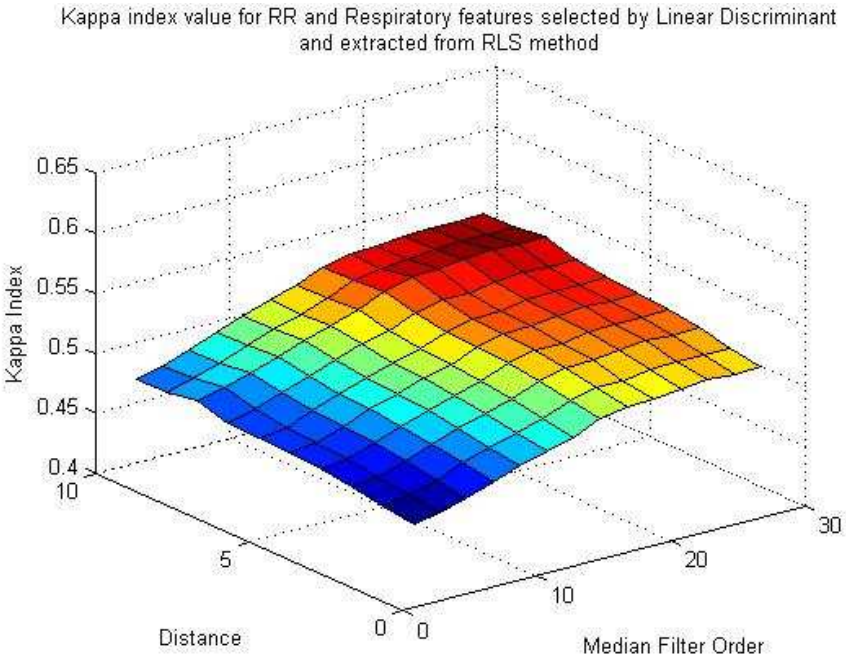


Figure 5.4: Median filter degree and optimal distances based on mean kappa index for each subject computing with LD and using RR and respiratory features extracted with RLS method. Median filter order=23 and distance=10 maximize mean kappa index value.

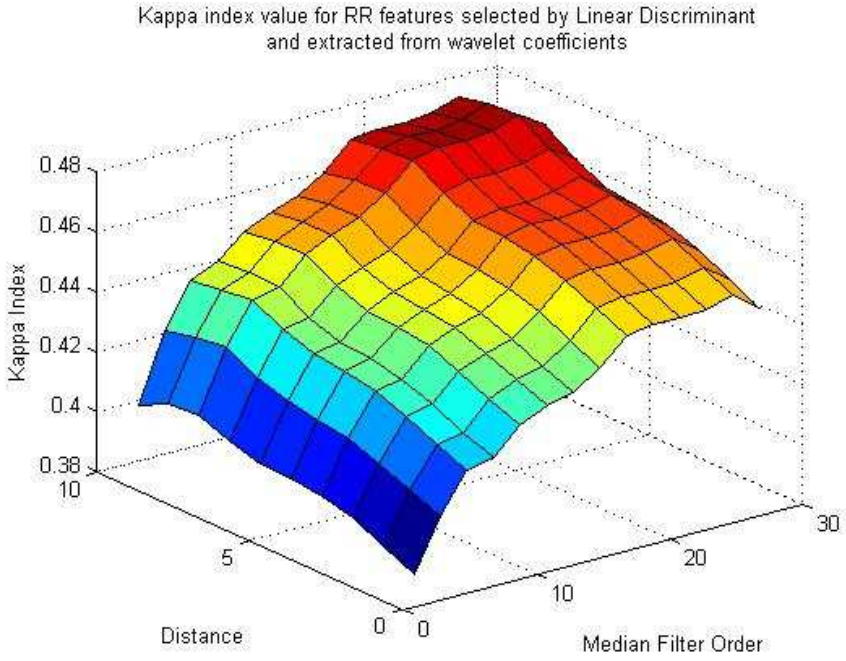


Figure 5.5: Median filter degree and optimal distances based on mean kappa index for each subject computing with LD and using RR features extracted with wavelet method. Median filter order=23 and distance=10 maximize mean kappa index value.

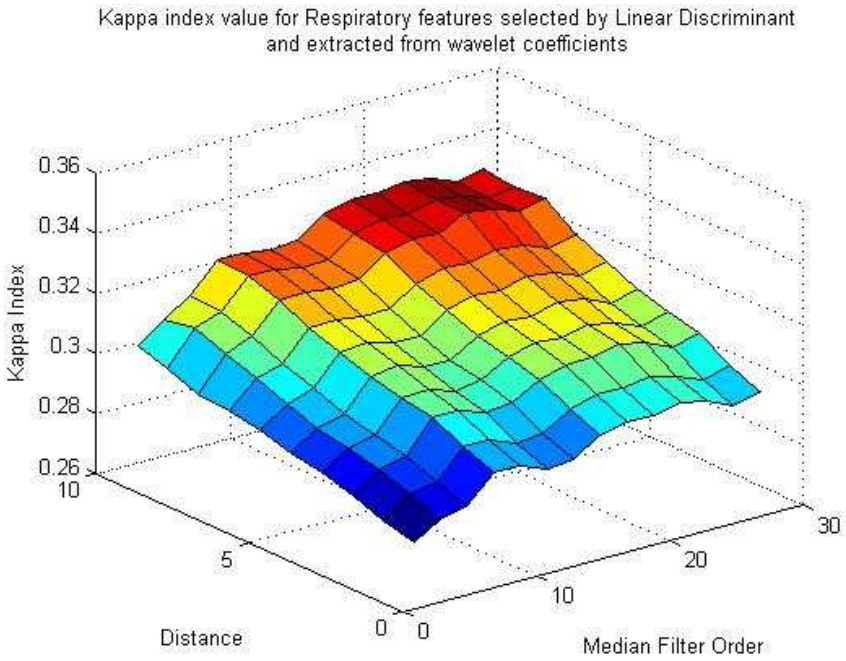


Figure 5.6: Median filter degree and optimal distances based on mean kappa index for each subject computing with LD and using respiratory features extracted with wavelet method. Median filter order=27 and distance=8 maximize mean kappa index value.

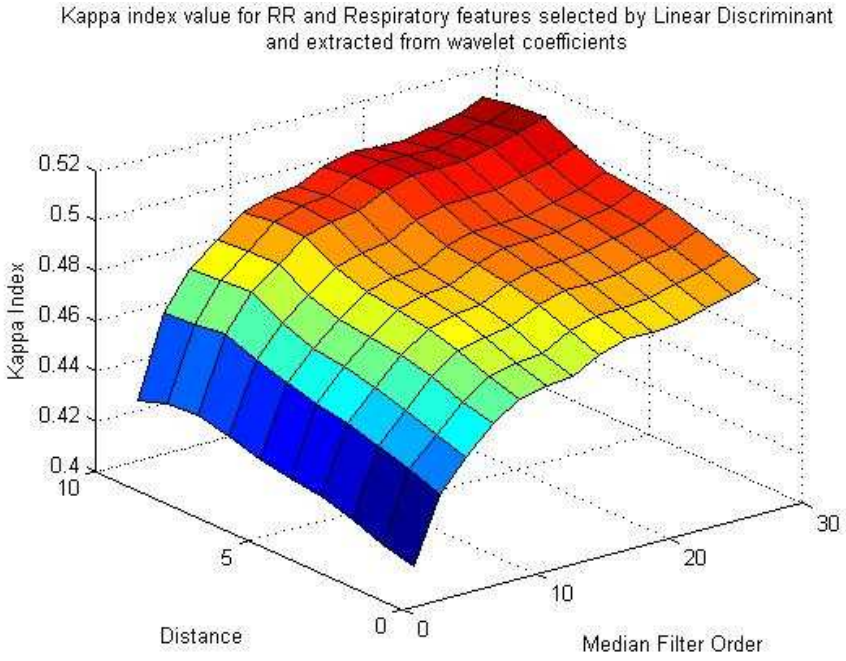


Figure 5.7: Median filter degree and optimal distances based on mean kappa index for each subject computing with LD and using RR and respiratory features extracted with wavelet method Median filter order=23 and distance=10 maximize mean kappa index value.

The following table (Table 5.1) summarizes results of degree and feature selection with LD and QD classifier. One can note as the first features selected from both classifiers are more statistically significant than last. Feature numbers correspond to features illustrated in subparagraph 4.5.

Signal used	**Feature extraction method**	**Feature selection method**	**Feature selected for each sleep stages**			**Median Filter Degree**
			WAKE	NREM	REM	
Only RR	RLS	LD	3,2,7,1	3,2,7,1	3,1,2	25
Only RR	RLS	QD	3,2,5,4	3,2	3,7,1,8,4	25
Only RR	Wavelet	LD	9,10,11,2,1,3	9,13,1,10	9,13,12,11,2,1	27
Only RR	Wavelet	QD	9,10,2,11,5,3	9,13,10,11	9,13,12,2,11,8	27
Only	RLS	LD	12,14	9,13,11	9,13,11	23

Respiratory						
Only Respiratory	RLS	QD	10	14,13,11,9, 12	14,10,11,12, 13,9	9
Only Respiratory	Wavelet	LD	20	20,21,24,1 4,19	21,24,22,19	23
Only Respiratory	Wavelet	QD	20	20,21,24,1 6,14,23	21,24,22,19	19
RR and Respiratory	RLS	LD	14,5,9,2	5,15,3,2,9, 13,1	5,15,3,2,1,6	25
RR and Respiratory	RLS	QD	10,5,11,2	3,9,2,7,15, 17,6,10,5, 1	3,10,11,7,2,1 2,8,4,14	27
RR and Respiratory	Wavelet	LD	20,9,10, 11,1,2,26	25,13,1,16	25,13,28,26, 11,1,8	29
RR and Respiratory	Wavelet	QD	20,1,2,26, 29,27,19, 7,11,12,5, 13,14, 18,16,3, 10,15,21	25,13,11, 10,24,19, 2,9,22,29, 12	25,13,21,12, 11,24	25

Table 5.1: Results of feature selection. Numbers of features correspond to the same indicated in chapter 4.5. Optimal order of median filter was computed as the order maximizing mean kappa index value for both extracting methods and for each combination. Selection criterion is based on both linear and quadratic discriminant analysis.

The following tables (table 5.2) show results obtained with LD for each subject. Combination of RR and respiratory features improves the performance of algorithm and it maximizes mean kappa index value when optimal median filter was applied (order=25 for RLS model and order=23 for wavelet decomposition) after imposing the epochs between 2 movement events, that occur in a certain interval (interval=8 epochs for RLS model and for wavelet decomposition), as WAKE stages. Accuracy and kappa index are used in order to compare results. Table 5.2a and 5.2d show single-stage accuracy and kappa index, 3-state accuracy and kappa index after merging for feature extracted with RLS and wavelet method respectively. Table 5.2b and 5.2e illustrates 2 state accuracy, considering one class

against all for each class after merging and table 5.2c and 5.2f compare medical and estimated sleep efficiency for both methods.

RLS based features:	WAKE		NREM		REM		TOTAL	
	Accuracy	Kappa index	Accuracy	Kappa index	Accuracy	Kappa index	Accuracy	Kappa index
RR	0,666	0,172	0,739	0,429	0,723	0,312	0,651	0,343
Respiratory	0,607	0,168	0,648	0,319	0,621	0,205	0,562	0,249
RR + Respiratory	0,656	0,210	0,765	0,494	0,744	0,357	0,675	0,387
RR + movement	0,675	0,225	0,747	0,454	0,745	0,341	0,675	0,398
Respiratory + movement	0,615	0,207	0,656	0,338	0,647	0,232	0,589	0,303
RR + Respiratory + movement	0,662	0,243	0,771	0,514	0,768	0,389	0,699	0,440
RR + movement + smooth	0,667	0,119	0,760	0,468	0,800	0,441	0,745	0,504
Respiratory + movement + smooth	0,594	0,129	0,652	0,335	0,667	0,268	0,626	0,354
RR + Respiratory + movement + smooth	0,649	0,139	0,778	0,521	0,819	0,483	0,764	0,536

Table 5.2a: Single state and 3-state accuracy are presented. The last column on the right shows mean kappa index value after merging. The best result is with RR and respiratory RLS based feature combination applying optimal median filter (order=23).

RLS based features:	WAKE vs NREM+REM		NREM vs WAKE+REM		REM vs WAKE+NREM	
	Accuracy	Kappa index	Accuracy	Kappa index	Accuracy	Kappa index
RR	0,828	0,127	0,745	0,446	0,730	0,318
Respiratory	0,841	0,135	0,649	0,322	0,635	0,222
RR + Respiratory	0,829	0,118	0,769	0,505	0,752	0,368
RR + movement	0,850	0,330	0,752	0,466	0,749	0,344
Respiratory + movement	0,863	0,363	0,656	0,341	0,657	0,245
RR + Respiratory + movement	0,851	0,323	0,775	0,522	0,773	0,396
RR + movement + smooth	0,888	0,483	0,797	0,541	0,804	0,450
Respiratory + movement + smooth	0,886	0,486	0,670	0,356	0,697	0,298
RR + Respiratory + movement + smooth	0,890	0,492	0,809	0,572	0,829	0,503

Table 5.2b: 2 states accuracy and kappa index for each class against all after merging. Features are extracted with RLS method.

RLS based features:	Sleep efficiency		REM during TST		1st REM Latency	
	R&K	Automatic	R&K	Automatic	R&K	Automatic
RR	0,852	0,935	0,201	0,350	86,441	22,088
Respiratory	0,852	0,954	0,201	0,436	86,441	13,824
RR + Respiratory	0,852	0,941	0,201	0,351	86,441	18,118
RR + movement	0,852	0,901	0,201	0,327	86,441	22,176
Respiratory + movement	0,852	0,919	0,201	0,410	86,441	14,000
RR + Respiratory + movement	0,852	0,907	0,201	0,326	86,441	18,206
RR + movement + smooth	0,852	0,921	0,201	0,278	86,441	85,706
Respiratory + movement + smooth	0,852	0,913	0,201	0,373	86,441	63,853
RR + Respiratory + movement + smooth	0,852	0,923	0,201	0,272	86,441	93,294

Table 5.2c: Medical and automatic sleep efficiency comparing.

Wavelet based features:	WAKE		NREM		REM		TOTALE	
	Accuracy	Kappa index	Accuracy	Kappa index	Accuracy	Kappa index	Accuracy	Kappa index
RR	0,698	0,217	0,714	0,382	0,672	0,252	0,625	0,304
Respiratory	0,648	0,193	0,632	0,239	0,603	0,177	0,556	0,209
RR + Respiratory	0,694	0,235	0,717	0,396	0,674	0,267	0,626	0,317
RR + movement	0,699	0,225	0,717	0,389	0,696	0,281	0,646	0,348
Respiratory + movement	0,649	0,195	0,634	0,244	0,620	0,193	0,571	0,240
RR + Respiratory + movement	0,695	0,239	0,719	0,403	0,701	0,298	0,647	0,361
RR + movement + smooth	0,706	0,091	0,763	0,470	0,760	0,406	0,728	0,477
Respiratory + movement + smooth	0,642	0,098	0,650	0,266	0,659	0,290	0,648	0,330
RR + Respiratory + movement + smooth	0,688	0,101	0,756	0,469	0,775	0,447	0,739	0,512

Table 5.2d: Single state and 3-state accuracy are presented. The last column on the right shows mean kappa index value after merging. The best result is with RR wavelet based feature combination applying optimal median filter (order=25).

Wavelet based features:	WAKE vs NREM+REM		NREM vs WAKE+REM		REM vs WAKE+NREM	
	Accuracy	Kappa index	Accuracy	Kappa index	Accuracy	Kappa index
RR	0,834	0,169	0,719	0,395	0,697	0,255
Respiratory	0,824	0,216	0,635	0,249	0,653	0,156
RR + Respiratory	0,829	0,159	0,722	0,409	0,701	0,278
RR + movement	0,852	0,337	0,721	0,400	0,719	0,283
Respiratory + movement	0,837	0,322	0,637	0,254	0,668	0,172
RR + Respiratory + movement	0,847	0,325	0,724	0,414	0,724	0,307
RR + movement + smooth	0,890	0,487	0,782	0,505	0,785	0,409
Respiratory + movement + smooth	0,884	0,484	0,690	0,322	0,721	0,232
RR + Respiratory + movement + smooth	0,889	0,484	0,790	0,537	0,799	0,471

Table 5.2e: Bi-state accuracy and kappa index, using as classes WAKE against all and NREM against all. Medical sleep efficiency is compared to estimated one. Features are extracted from wavelet coefficients.

Wavelet based features:	Sleep efficiency		REM during TST		1st REM Latency	
	R&K	Automatic	R&K	Automatic	R&K	Automatic
RR	0,852	0,931	0,201	0,352	86,441	14,147
Respiratory	0,852	0,898	0,201	0,357	86,441	9,265
RR + Respiratory	0,852	0,927	0,201	0,366	86,441	12,294
RR + movement	0,852	0,903	0,201	0,326	86,441	14,588
Respiratory + movement	0,852	0,878	0,201	0,340	86,441	9,412
RR + Respiratory + movement	0,852	0,897	0,201	0,339	86,441	14,382
RR + movement + smooth	0,852	0,923	0,201	0,279	86,441	74,647
Respiratory + movement + smooth	0,852	0,909	0,201	0,284	86,441	102,912
RR + Respiratory + movement + smooth	0,852	0,923	0,201	0,291	86,441	83,824

Table 5.2f: Medical and automatic sleep efficiency comparing.

Figure 5.8 shows an example of automatic sleep staging using RR and Respiratory features extracted from RLS model (middle box) parameters and wavelet coefficients (bottom box) in according to R&K methods (upper box).

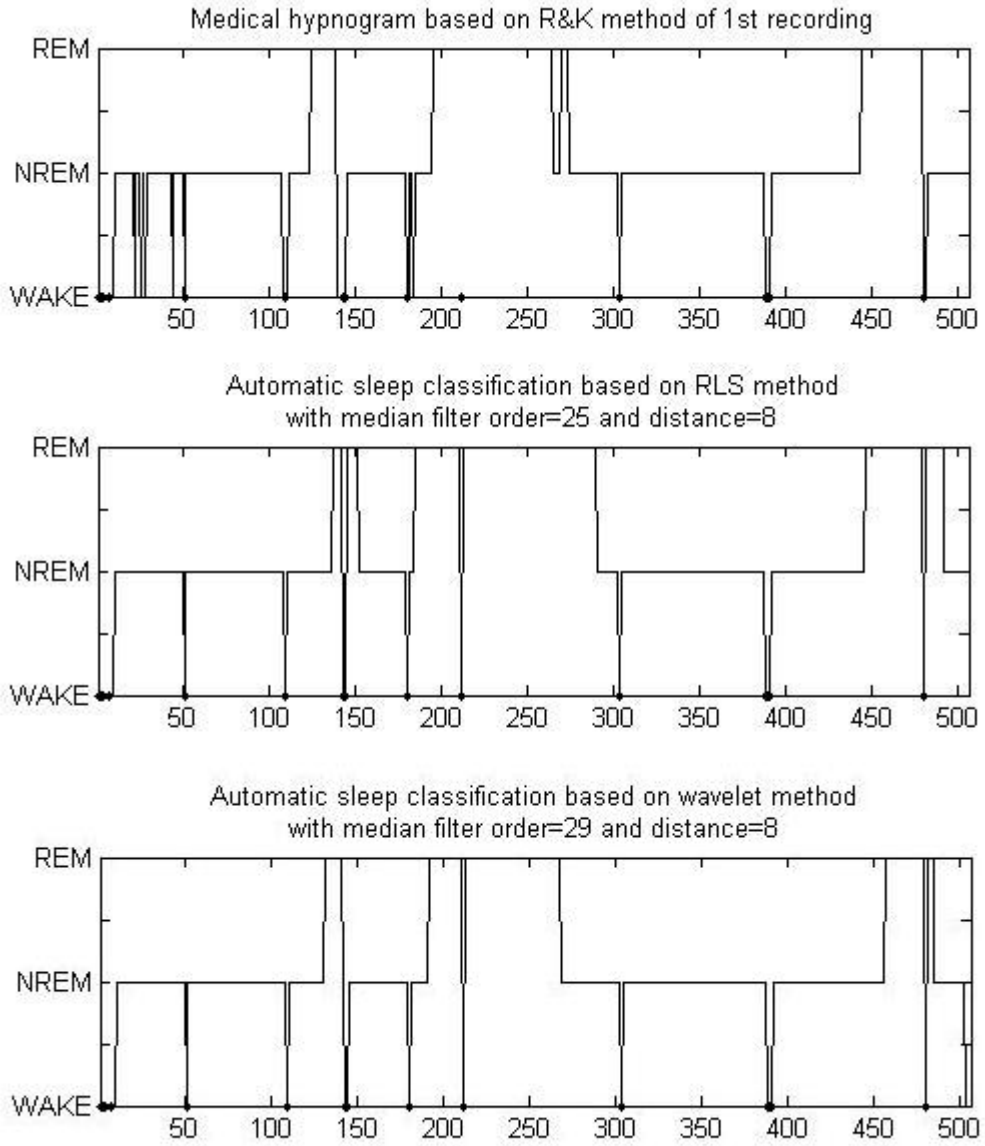


Figure 5.8: The upper plot shows medical hypnogram according to R&K method. In the middle plot, an example of automatic hypnogram using RLS and wavelet (below) based LD feature selection.

5.2 Quadratic Discriminant

Results obtained with the quadratic discriminant classifier are shown in tables 5.3a/b/c (analysis on features obtained with RLS model) and in tables 5.3d/e/f (analysis on features obtained with Wavelet decomposition). Used features are chosen by QD analysis. The

following figures represent the grid search of optimal median filter order from 1 to 29 and optimal distances between 2 movement events (interval=8 epochs for RLS model and for wavelet decomposition). The best couple maximizes curve determined as mean kappa index value on each subject.

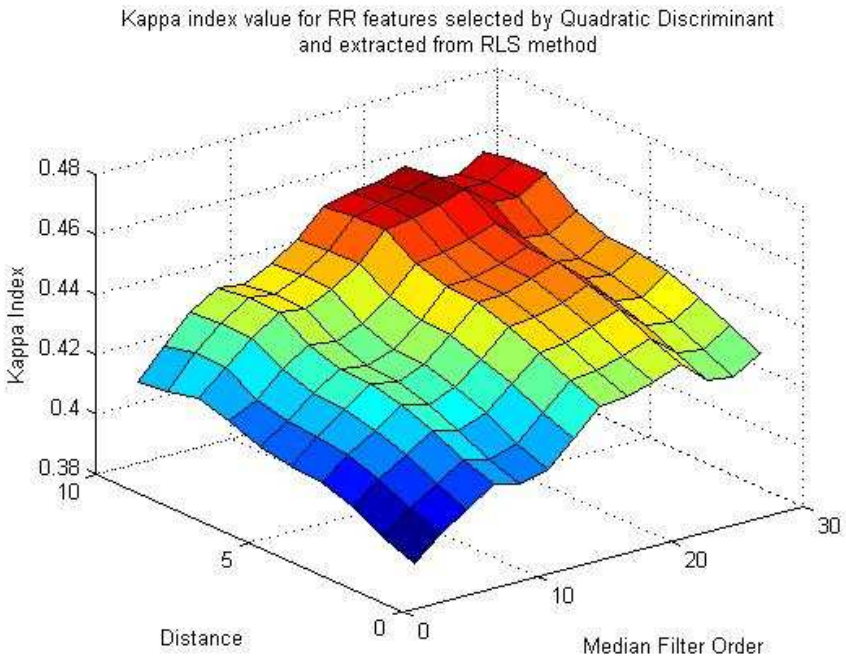


Figure 5.9: : Median filter degree and optimal distances based on mean kappa index for each subject computing with QD and using RR features extracted with RLS method. Median filter order=23 and distance=8 maximize mean kappa index value.

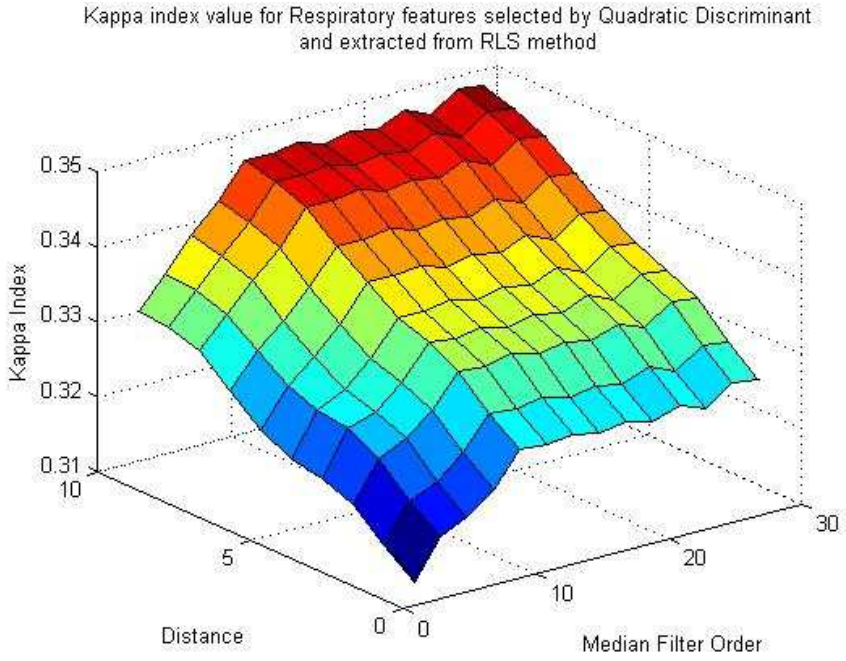


Figure 5.10: Median filter degree and optimal distances based on mean kappa index for each subject computing with QD and using respiratory features extracted with RLS method. Median filter order=27 and distance=10 maximize mean kappa index value.

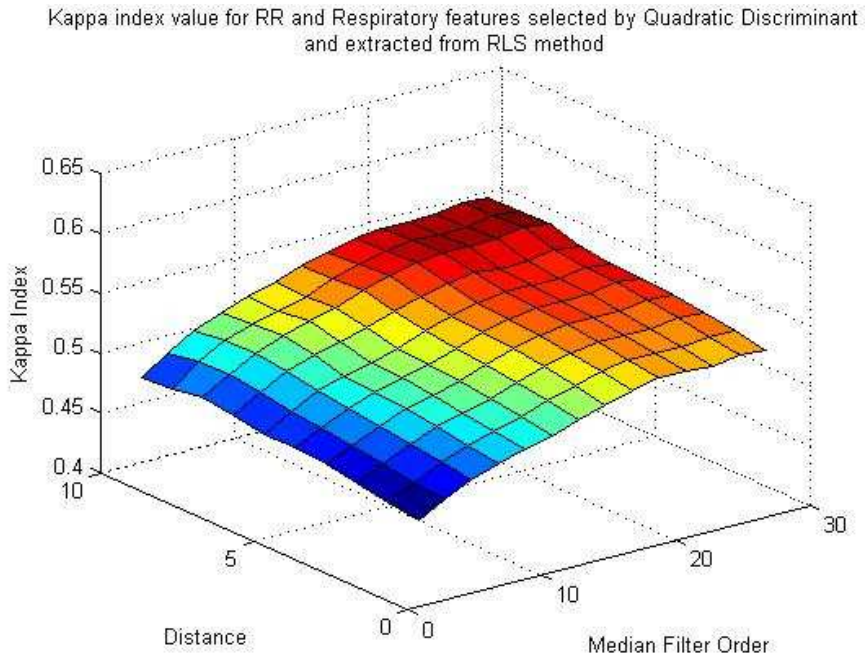


Figure 5.11: Median filter degree and optimal distances based on mean kappa index for each subject computing with QD and using RR and respiratory features extracted with RLS method. Median filter order=27 and distance=8 maximize mean kappa index value.

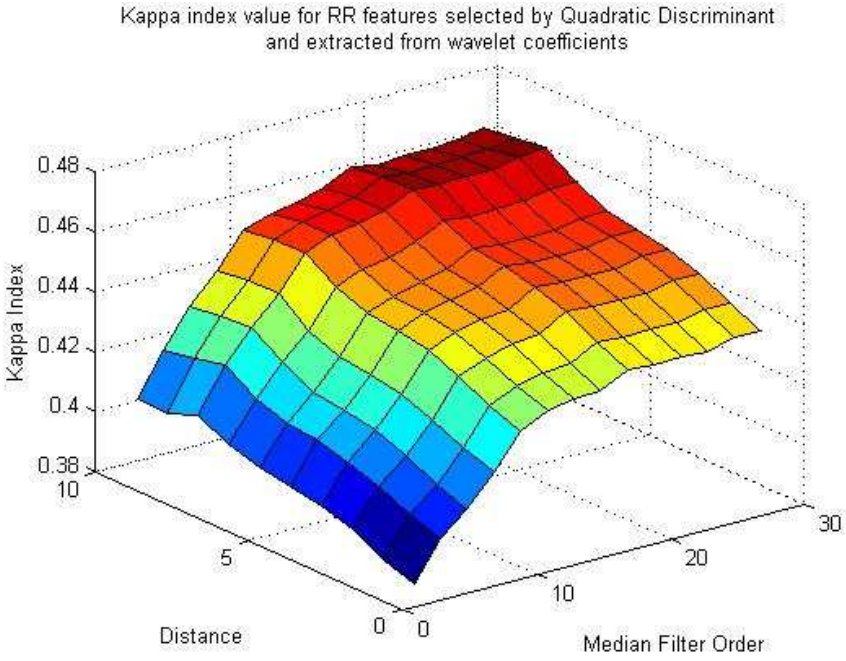


Figure 5.12: Median filter degree and optimal distances based on mean kappa index for each subject computing with QD using RR features extracted with wavelet method. Median filter order=27 and distance=8 maximize mean kappa index value.

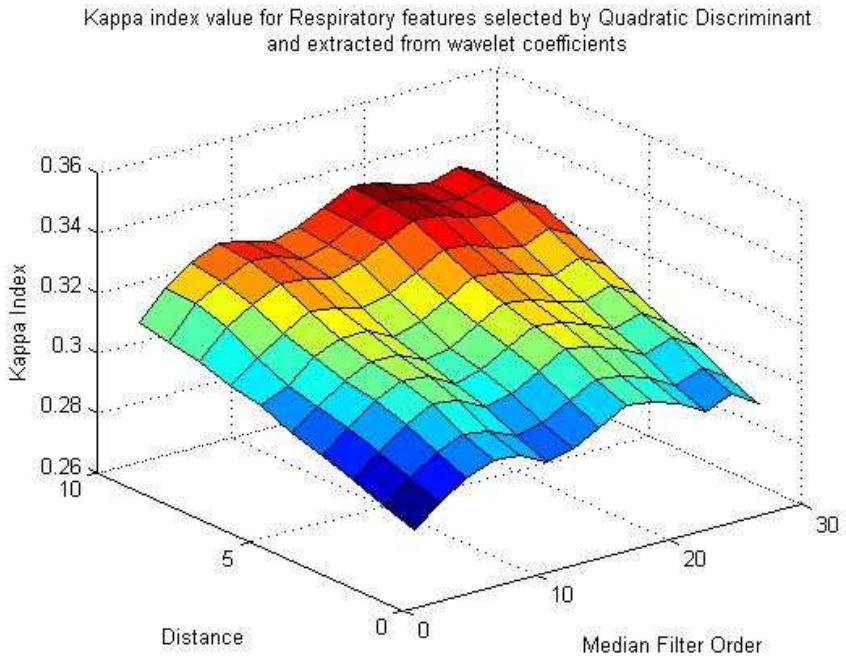


Figure 5.13 Median filter degree and optimal distances based on mean kappa index for each subject computing with QD using respiratory features extracted with wavelet method. Median filter order=19 and distance=10 maximize mean kappa index value.

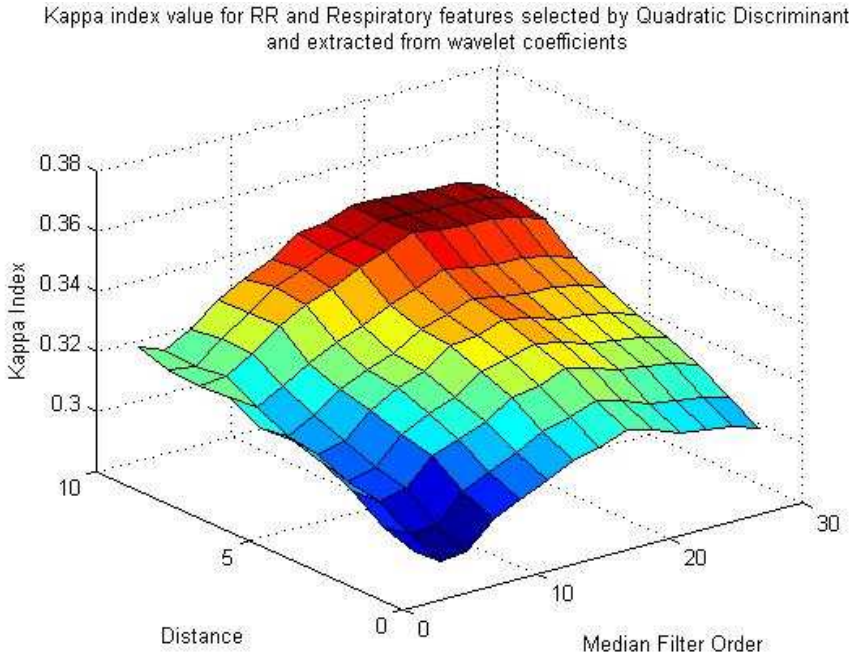


Figure 5.14 Median filter degree and optimal distances based on mean kappa index for each subject computing with QD using RR and respiratory features extracted with wavelet method. Median filter order=19 and distance=8 maximize mean kappa index value.

RLS based features:	WAKE		NREM		REM		TOTAL	
	Accuracy	Kappa index	Accuracy	Kappa index	Accuracy	Kappa index	Accuracy	Kappa index
RR	0,624	0,149	0,724	0,397	0,694	0,282	0,635	0,321
Respiratory	0,634	0,169	0,631	0,315	0,576	0,221	0,543	0,257
RR + Respiratory	0,695	0,239	0,763	0,493	0,743	0,365	0,675	0,395
RR + movement	0,632	0,201	0,732	0,423	0,716	0,309	0,660	0,378
Respiratory + movement	0,642	0,215	0,637	0,330	0,604	0,247	0,568	0,306
RR + Respiratory + movement	0,703	0,291	0,769	0,510	0,765	0,395	0,697	0,445
RR + movement + smooth	0,624	0,102	0,741	0,427	0,742	0,359	0,706	0,458
Respiratory + movement + smooth	0,617	0,107	0,629	0,325	0,609	0,285	0,590	0,349
RR + Respiratory + movement + smooth	0,687	0,150	0,776	0,521	0,820	0,495	0,768	0,550

Table 5.3a: Single state and 3-state accuracy are presented. The last column on the right shows mean kappa index value after merging. The best result is with RR and respiratory RLS based feature combination applying optimal median filter (order=25).

RLS based features:	WAKE vs NREM+REM		NREM vs WAKE+REM		REM vs WAKE+NREM	
	Accuracy	Kappa index	Accuracy	Kappa index	Accuracy	Kappa index
RR	0,837	0,138	0,730	0,415	0,704	0,292
Respiratory	0,851	0,175	0,632	0,318	0,603	0,239
RR + Respiratory	0,823	0,148	0,766	0,502	0,760	0,382
RR + movement	0,859	0,354	0,737	0,436	0,723	0,316
Respiratory + movement	0,872	0,405	0,638	0,332	0,627	0,262
RR + Respiratory + movement	0,844	0,340	0,772	0,518	0,779	0,409
RR + movement + smooth	0,888	0,480	0,768	0,490	0,757	0,383
Respiratory + movement + smooth	0,876	0,467	0,643	0,352	0,660	0,316
RR + Respiratory + movement + smooth	0,886	0,495	0,813	0,584	0,837	0,517

Table 5.3b: 2 states accuracy and kappa index for each class against all after merging.

Features are extracted with RLS method.

RLS based features:	Sleep efficiency		REM during TST		1st REM Latency	
	R&K	Automatic	R&K	Automatic	R&K	Automatic
RR	0,852	0,948	0,201	0,378	86,441	24,676
Respiratory	0,852	0,959	0,201	0,499	86,441	10,265
RR + Respiratory	0,852	0,914	0,201	0,346	86,441	19,059
RR + movement	0,852	0,913	0,201	0,354	86,441	26,029
Respiratory + movement	0,852	0,924	0,201	0,472	86,441	11,441
RR + Respiratory + movement	0,852	0,881	0,201	0,322	86,441	19,147
RR + movement + smooth	0,852	0,921	0,201	0,322	86,441	76,559
Respiratory + movement + smooth	0,852	0,900	0,201	0,439	86,441	57,853
RR + Respiratory + movement + smooth	0,852	0,903	0,201	0,269	86,441	92,324

Table 5.3c: Medical and automatic sleep efficiency comparing.

Wavelet based features:	WAKE		NREM		REM		TOTALE	
	Accuracy	Kappa index	Accuracy	Kappa index	Accuracy	Kappa index	Accuracy	Kappa index
RR	0,714	0,227	0,713	0,375	0,665	0,250	0,627	0,305
Respiratory	0,704	0,218	0,642	0,250	0,585	0,168	0,569	0,218
RR + Respiratory	0,676	0,160	0,687	0,305	0,628	0,152	0,600	0,233
RR + movement	0,715	0,237	0,715	0,382	0,690	0,279	0,647	0,347
Respiratory + movement	0,705	0,223	0,644	0,257	0,603	0,185	0,585	0,251
RR + Respiratory + movement	0,681	0,189	0,689	0,309	0,656	0,183	0,623	0,282
RR + movement + smooth	0,724	0,064	0,758	0,448	0,733	0,373	0,720	0,462
Respiratory + movement + smooth	0,694	0,074	0,657	0,272	0,638	0,277	0,648	0,333
RR + Respiratory + movement + smooth	0,725	0,160	0,695	0,285	0,719	0,254	0,693	0,347

Table 5.3d: Single state and 3-state accuracy are presented. The last column on the right shows mean kappa index value after merging. The best result is with RR and respiratory wavelet based feature combination applying optimal median filter (order=23).

Wavelet based features:	WAKE vs NREM+REM		NREM vs WAKE+REM		REM vs WAKE+NREM	
	Accuracy	Kappa index	Accuracy	Kappa index	Accuracy	Kappa index
RR	0,837	0,175	0,718	0,389	0,699	0,258
Respiratory	0,836	0,221	0,645	0,261	0,657	0,164
RR + Respiratory	0,812	0,144	0,693	0,320	0,696	0,160
RR + movement	0,854	0,341	0,720	0,393	0,721	0,286
Respiratory + movement	0,850	0,341	0,647	0,266	0,673	0,182
RR + Respiratory + movement	0,832	0,317	0,693	0,322	0,721	0,194
RR + movement + smooth	0,889	0,482	0,772	0,483	0,779	0,395
Respiratory + movement + smooth	0,884	0,481	0,690	0,323	0,723	0,248
RR + Respiratory + movement + smooth	0,884	0,471	0,730	0,349	0,772	0,231

Table 5.3e: Bi-state accuracy and kappa index, using as classes WAKE against all and NREM against all. Medical sleep efficiency is compared to estimated. Features are extracted from wavelet coefficients.

Wavelet based features:	Sleep efficiency		REM during TST		1st REM Latency	
	R&K	Automatic	R&K	Automatic	R&K	Automatic
RR	0,852	0,934	0,201	0,352	86,441	17,147
Respiratory	0,852	0,917	0,201	0,355	86,441	8,588
RR + Respiratory	0,852	0,902	0,201	0,286	86,441	15,735
RR + movement	0,852	0,905	0,201	0,326	86,441	17,647
Respiratory + movement	0,852	0,896	0,201	0,337	86,441	9,000
RR + Respiratory + movement	0,852	0,871	0,201	0,257	86,441	16,559
RR + movement + smooth	0,852	0,922	0,201	0,280	86,441	74,765
Respiratory + movement + smooth	0,852	0,910	0,201	0,287	86,441	104,353
RR + Respiratory + movement + smooth	0,852	0,917	0,201	0,178	86,441	110,265

Table 5.3f: Medical and automatic sleep efficiency comparing.

Combination of RR and respiratory features improves the performance of RLS algorithm and it maximizes mean kappa index value when optimal median filter was applied (order=23); while the features extracted by only RR signals extracted by wavelet coefficients improves the performance of classification, applying optimal median filter (order=25). Accuracy and kappa index are used in order to compare results. Table 5.3a and 5.3d show single-stage accuracy and kappa index, 3-state accuracy and kappa index after merging for feature extracted with RLS and wavelet method respectively. Table 5.3b and 5.3e illustrates 2 state accuracy, considering one class against all for each class after merging and table 5.3c and 5.3f compare medical and estimated sleep efficiency for both methods. Figure 5.15 shows an example of automatic sleep staging using RR and Respiratory features extracted from RLS model (middle box) parameters and wavelet coefficients (bottom box) in according to R&K methods (upper box).

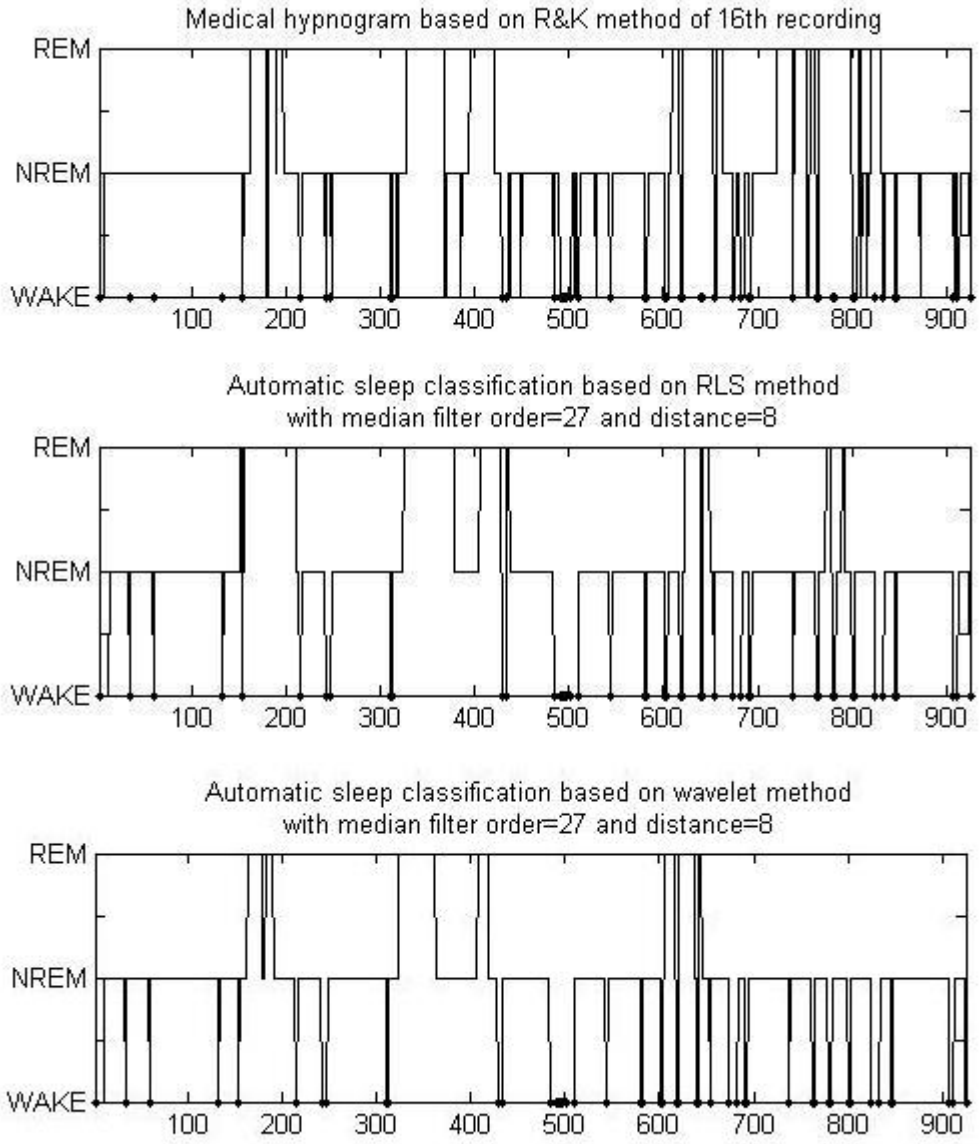


Figure 5.15: The upper plot shows medical hypnogram according to R&K method. In the middle plot, an example of automatic hypnogram using RLS and wavelet (below) based QD feature selection.

5.3 *K*-Nearest-Neighbor

K-Nearest-Neigbor uses features selected with LD analysis. A search of the optimal K was done using odd values of $K = 1–29$ in data sets with one class against all for each sleep stage. Figure 5.16 shows the best combination of *K* and median filter order applied to the

automatic sleep staging for every features extracted with both methods (RLS model and wavelet decomposition). The best result maximizes curve determined as mean kappa index value for all subject. K-NN research the distance between 2 movements events to Movement activity and median filter using were compared each one. It was selected the optimal distances between 2 movement events and median filter degree selecting iteratively starting from 1 to 10 and from 1 to 29, respectively, as maximizing mean kappa index value, as shown in figures from 5.16 to 5.21, which showing mean kappa index.

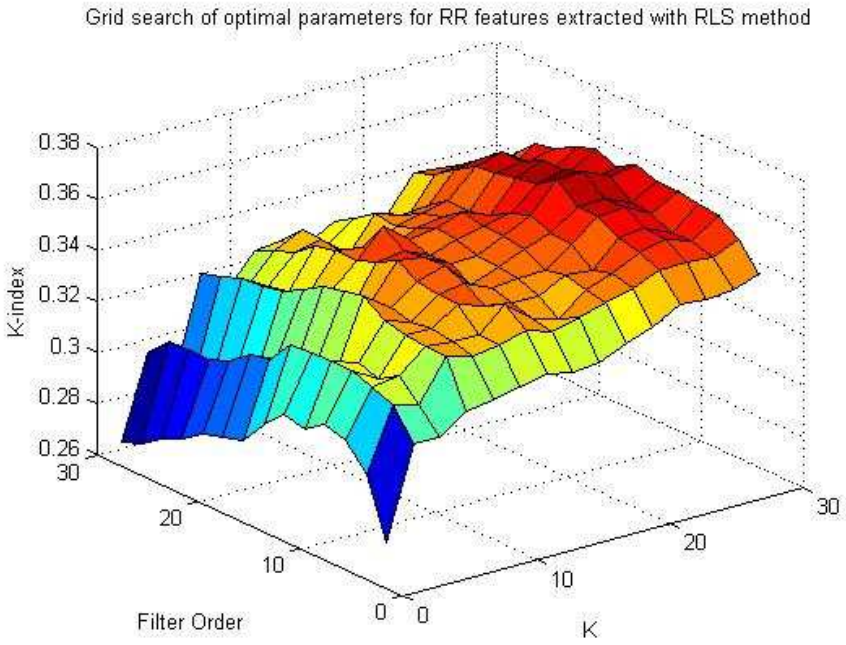


Figure 5.16: Grid search of optimal parameters for K-NN classifier on RR features extracted with RLS method analysis. K=25 and order=15 maximize mean kappa index value with WAKE interval=10.

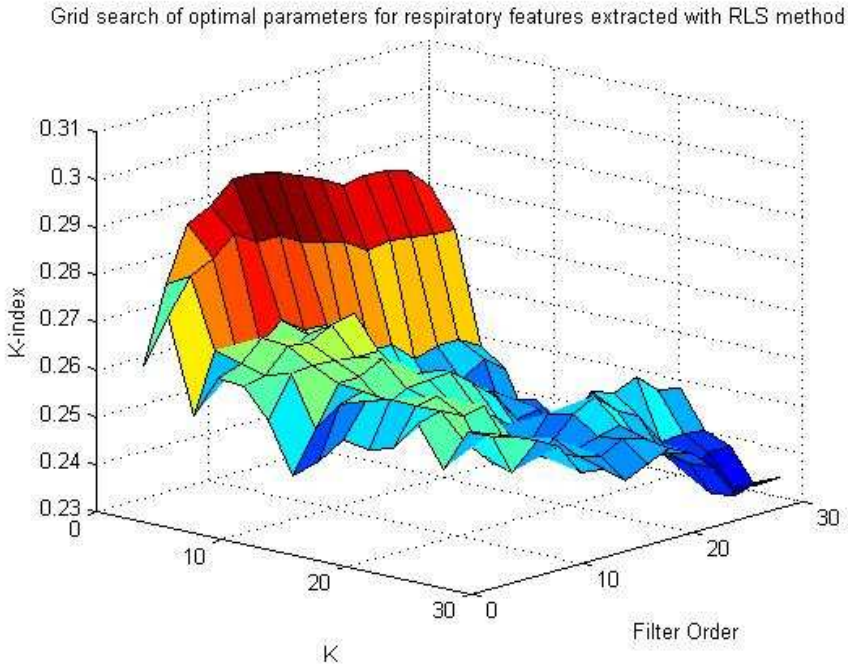


Figure 5.17: Grid search of optimal parameters for K-NN classifier on respiratory features extracted with RLS method analysis. K=1 and order=15 maximize mean kappa index value with WAKE interval=9.

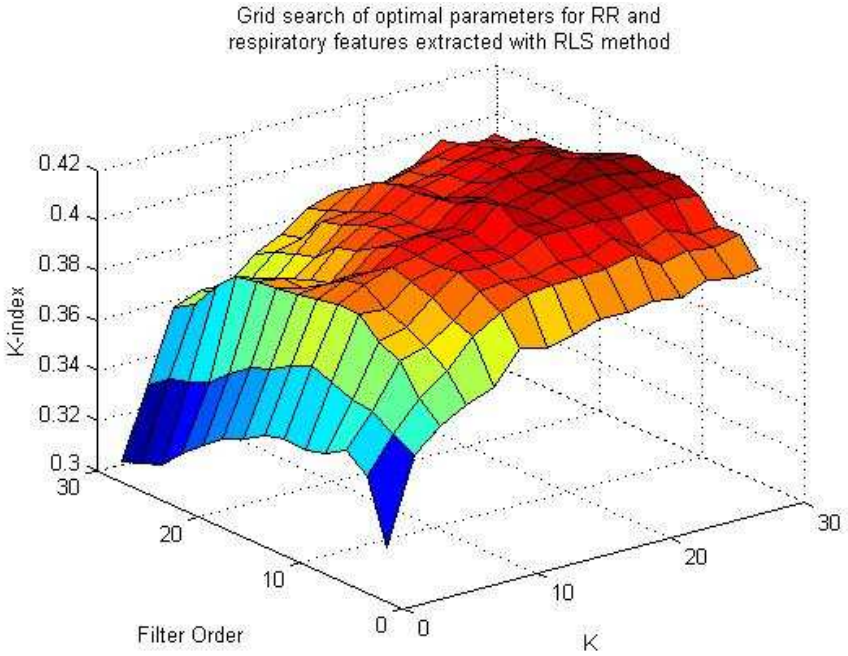


Figure 5.18: Grid search of optimal parameters for K-NN classifier on RR and respiratory features extracted with RLS method analysis. K=25 and order=15 maximize mean kappa index value with WAKE interval=10.

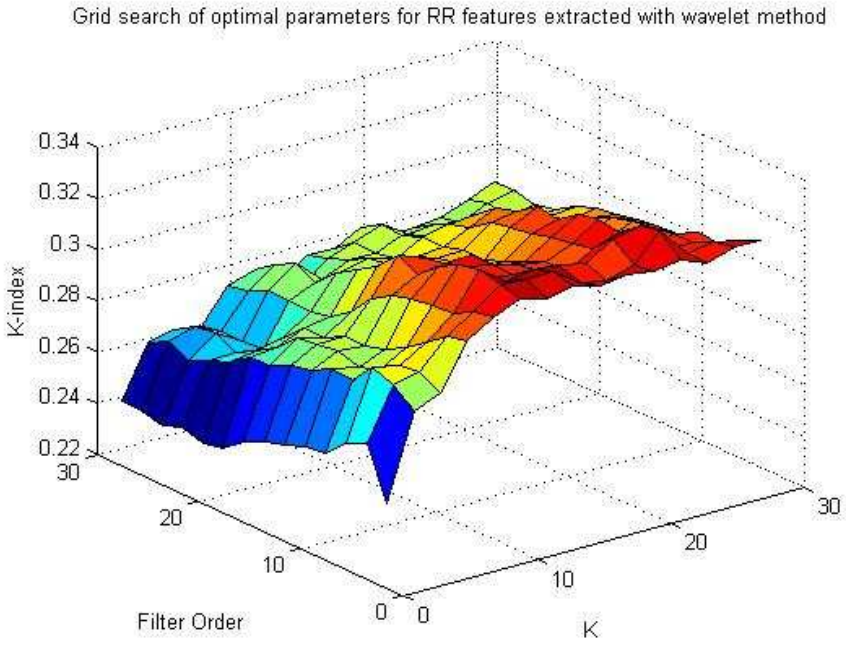


Figure 5.19: Grid search of optimal parameters for K-NN classifier on RR features extracted with wavelet method analysis. K=15 and order=5 maximize mean kappa index value with WAKE interval=9.

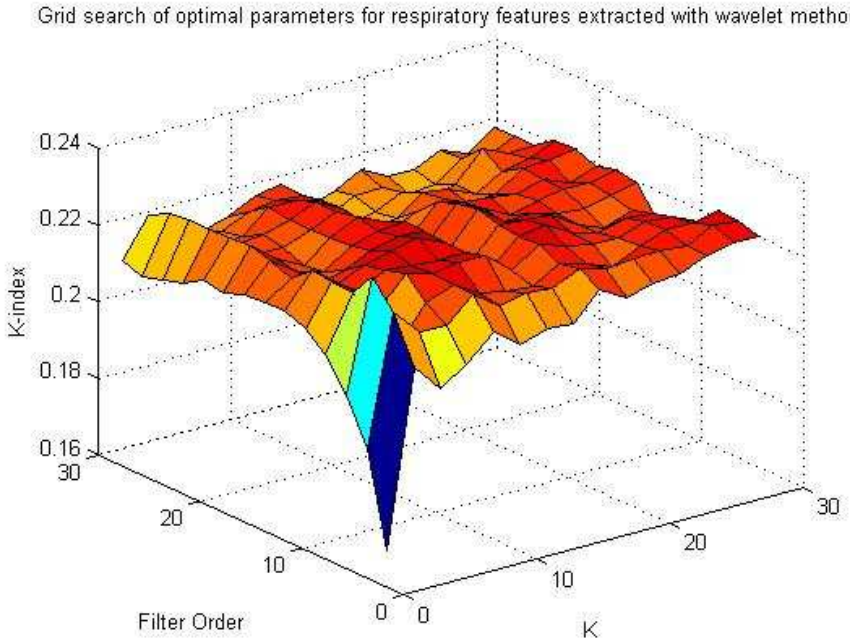


Figure 5.20: Grid search of optimal parameters for K-NN classifier on respiratory features extracted with wavelet method analysis. K=3 and order=7 maximize mean kappa index value with WAKE interval=9.

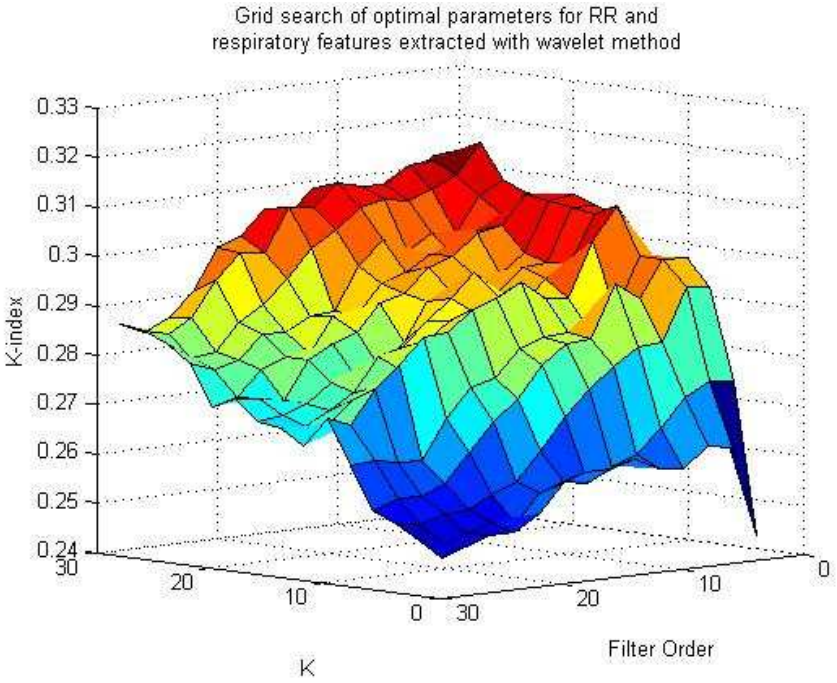


Figure 5.21: Grid search of optimal parameters for K-NN classifier on RR and respiratory features extracted with wavelet method analysis. K=25 and order=3 maximize mean kappa index value with WAKE interval=9.

Results obtained are shown in tables 5.4a/b/c (analysis on features obtained with RLS model) and in tables 5.4d/e/f (analysis on features obtained with Wavelet decomposition). Due to a mean kappa index value of WAKE better than REM for features extracted from wavelet coefficients, 3-state merging is based on two bi-state classifications: NREM versus all (REM and WAKE stages) detecting epochs with NREM stage; WAKE versus REM discriminating WAKE epochs and REM stages when epochs are not classified; while for features extracted from RLS method the classification is the same explained in chapter 4.5. The best result is obtained with RR and respiratory features extracted by RLS model, using movement for WAKE stages identification, *K*=25 and best median filter order=15 and with RR features extracted by wavelet coefficients, with movement identification, *K*=15 and median filter order =5, as shown in the previous figures. Figure 5.22 shows an example of automatic sleep staging using RR and Respiratory features extracted from RLS model (middle box) parameters and wavelet coefficients (bottom box) in according to R&K methods (upper box).

RLS based features:	WAKE		NREM		REM		TOTAL	
	Accuracy	Kappa index	Accuracy	Kappa index	Accuracy	Kappa index	Accuracy	Kappa index
RR	0,849	0,025	0,736	0,387	0,788	0,138	0,641	0,279
Respiratory	0,759	0,057	0,611	0,163	0,714	0,125	0,546	0,145
RR + Respiratory	0,843	0,054	0,762	0,451	0,803	0,226	0,670	0,338
RR + movement	0,875	0,351	0,747	0,424	0,794	0,148	0,655	0,322
Respiratory + movement	0,780	0,224	0,627	0,208	0,723	0,135	0,567	0,201
RR + Respiratory + movement	0,867	0,338	0,774	0,485	0,809	0,237	0,685	0,379
RR + movement + smooth	0,853	0,020	0,757	0,433	0,817	0,154	0,693	0,372
Respiratory + movement + smooth	0,845	0,121	0,685	0,304	0,806	0,147	0,696	0,316
RR + Respiratory + movement + smooth	0,850	0,034	0,789	0,506	0,832	0,241	0,720	0,420

Table 5.4a: Single state and 3-state accuracy are presented. The last column on the right shows mean kappa index value after merging. The best result is with RR and respiratory RLS based feature combination applying optimal median filter (order=15).

RLS based features:	WAKE vs NREM+REM		NREM vs WAKE+REM		REM vs WAKE+NREM	
	Accuracy	Kappa index	Accuracy	Kappa index	Accuracy	Kappa index
RR	0,748	0,192	0,743	0,407	0,790	0,141
Respiratory	0,760	0,107	0,615	0,174	0,718	0,131
RR + Respiratory	0,768	0,217	0,768	0,468	0,805	0,229
RR + movement	0,762	0,279	0,752	0,438	0,796	0,150
Respiratory + movement	0,778	0,245	0,630	0,217	0,726	0,140
RR + Respiratory + movement	0,782	0,305	0,778	0,497	0,811	0,240
RR + movement + smooth	0,785	0,347	0,784	0,495	0,818	0,154
Respiratory + movement + smooth	0,865	0,467	0,720	0,316	0,807	0,159
RR + Respiratory + movement + smooth	0,805	0,374	0,800	0,531	0,834	0,238

Table 5.4b: 2 states accuracy and kappa index for each class against all after merging. Features are extracted with RLS method.

RLS based features:	Sleep efficiency		REM during TST		1st REM Latency	
	R&K	Automatic	R&K	Automatic	R&K	Automatic
RR	0,852	0,773	0,201	0,092	86,441	34,912
Respiratory	0,852	0,831	0,201	0,205	86,441	9,176
RR + Respiratory	0,852	0,798	0,201	0,109	86,441	30,794
RR + movement	0,852	0,750	0,201	0,085	86,441	34,971
Respiratory + movement	0,852	0,801	0,201	0,195	86,441	9,206
RR + Respiratory + movement	0,852	0,776	0,201	0,102	86,441	30,941
RR + movement + smooth	0,852	0,759	0,201	0,046	86,441	231,971
Respiratory + movement + smooth	0,852	0,877	0,201	0,058	86,441	155,735
RR + Respiratory + movement + smooth	0,852	0,788	0,201	0,056	86,441	207,824

Table 5.4c: Medical and automatic sleep efficiency comparing.

Wavelet based features:	WAKE		NREM		REM		TOTALE	
	Accuracy	Kappa index	Accuracy	Kappa index	Accuracy	Kappa index	Accuracy	Kappa index
RR	0,854	0,141	0,721	0,345	0,783	0,129	0,641	0,266
Respiratory	0,826	0,134	0,633	0,158	0,736	0,064	0,571	0,140
RR + Respiratory	0,858	0,175	0,736	0,376	0,792	0,091	0,647	0,275
RR + movement	0,872	0,351	0,726	0,361	0,785	0,132	0,646	0,284
Respiratory + movement	0,841	0,283	0,642	0,186	0,741	0,069	0,581	0,171
RR + Respiratory + movement	0,873	0,356	0,741	0,390	0,795	0,096	0,653	0,292
RR + movement + smooth	0,861	0,155	0,750	0,400	0,804	0,098	0,678	0,338
Respiratory + movement + smooth	0,858	0,141	0,684	0,224	0,799	0,036	0,655	0,240
RR + Respiratory + movement + smooth	0,869	0,265	0,754	0,410	0,802	0,060	0,647	0,328

Table 5.4d: Single state and 3-state accuracy are presented. The last column on the right shows mean kappa index value after merging. The best result is with RR wavelet based feature combination applying optimal median filter (order=25).

Wavelet based features:	WAKE vs NREM+REM		NREM vs WAKE+REM		REM vs WAKE+NREM	
	Accuracy	Kappa index	Accuracy	Kappa index	Accuracy	Kappa index
RR	0,764	0,238	0,728	0,366	0,789	0,127
Respiratory	0,727	0,162	0,639	0,176	0,777	0,044
RR + Respiratory	0,757	0,253	0,743	0,395	0,795	0,086
RR + movement	0,769	0,274	0,732	0,378	0,792	0,131
Respiratory + movement	0,735	0,216	0,647	0,201	0,781	0,049
RR + Respiratory + movement	0,762	0,285	0,746	0,407	0,798	0,090
RR + movement + smooth	0,780	0,352	0,772	0,461	0,804	0,077
Respiratory + movement + smooth	0,805	0,355	0,706	0,279	0,800	0,011
RR + Respiratory + movement + smooth	0,718	0,297	0,778	0,500	0,799	0,013

Table 5.4e: Bi-state accuracy and kappa index, using as classes WAKE against all and NREM against all. Medical sleep efficiency is compared to estimated. Features are extracted from wavelet coefficients.

Wavelet based features:	Sleep efficiency		REM during TST		1st REM Latency	
	R&K	Automatic	R&K	Automatic	R&K	Automatic
RR	0,852	0,775	0,201	0,071	86,441	31,559
Respiratory	0,852	0,751	0,201	0,056	86,441	28,324
RR + Respiratory	0,852	0,750	0,201	0,040	86,441	42,647
RR + movement	0,852	0,765	0,201	0,068	86,441	38,853
Respiratory + movement	0,852	0,736	0,201	0,052	86,441	28,382
RR + Respiratory + movement	0,852	0,741	0,201	0,038	86,441	42,676
RR + movement + smooth	0,852	0,736	0,201	0,019	86,441	190,000
Respiratory + movement + smooth	0,852	0,798	0,201	0,002	86,441	291,912
RR + Respiratory + movement + smooth	0,852	0,646	0,201	0,004	86,441	246,765

Table 5.4f: Medical and automatic sleep efficiency comparing

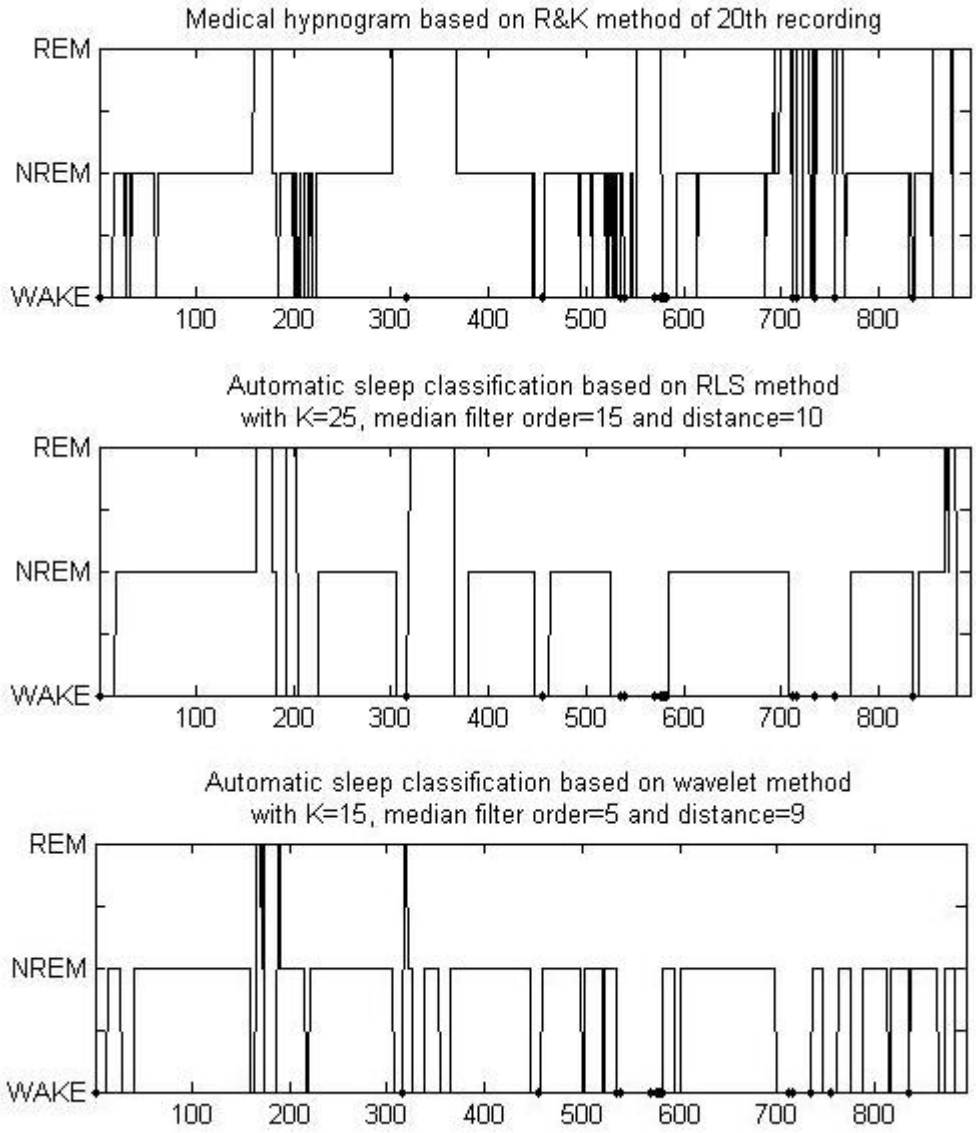


Figure 5.22: The upper plot shows medical hypnogram according to R&K method. In the middle plot, an example of automatic hypnogram using RLS and wavelet (below) based LD feature selection.

5.4 Neural Network

The architecture of Feed-Forward Neural Network was formed in three layers: input layer, hidden layer, and output layer. First layer had a number of neurons as features subset size and the output layer had two neurons selecting class. In order to the network to learn the desired mapping (neurons number of hidden layer) and at the same time avoid overfitting,

(neuron weights), early stopping and Leave One Out technique were used. Training function was implemented with the Levenberg-Marquardt back-propagation algorithm, updating network weights. Training occurs 15 times by starting with a random set of weights to avoid local minima. It was studied three different architectures for each sleep stages, three for only HRV features, three for only respiratory features and three for both. The best neurons number of hidden layer was computed considering the number of hidden layer's neurons which maximizes mean performances over all recordings. To avoid misclassification, three classes are shrewdly been balanced.

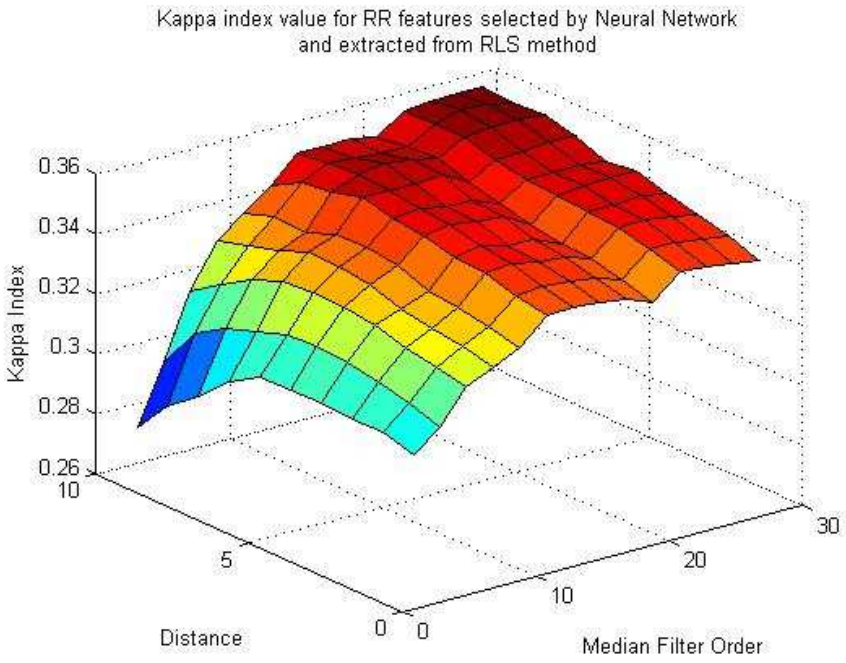


Figure 5.23: Median filter degree and optimal distances based on mean kappa index for each subject computing with FFNN and using RR features extracted with RLS method. Median filter order=25 and distance=9 maximize mean kappa index value.

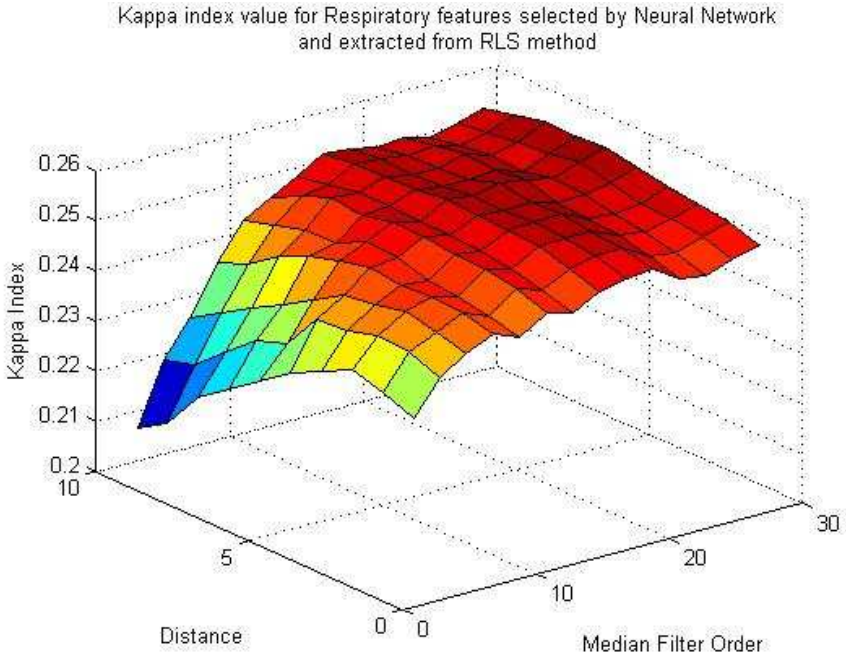


Figure 5.24: Median filter degree and optimal distances based on mean kappa index for each subject computing with FFNN and using respiratory features extracted with RLS method. Median filter order=29 and distance=6 maximize mean kappa index value.

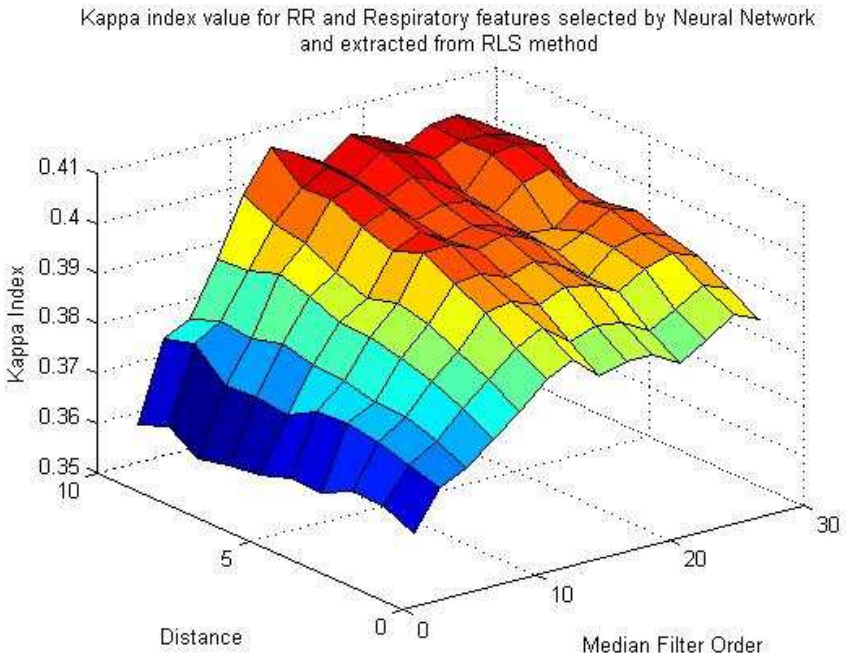


Figure 5.25: Median filter degree and optimal distances based on mean kappa index for each subject computing with FFNN and using RR and respiratory features extracted with RLS method. Median filter order=13 and distance=10 maximize mean kappa index value.

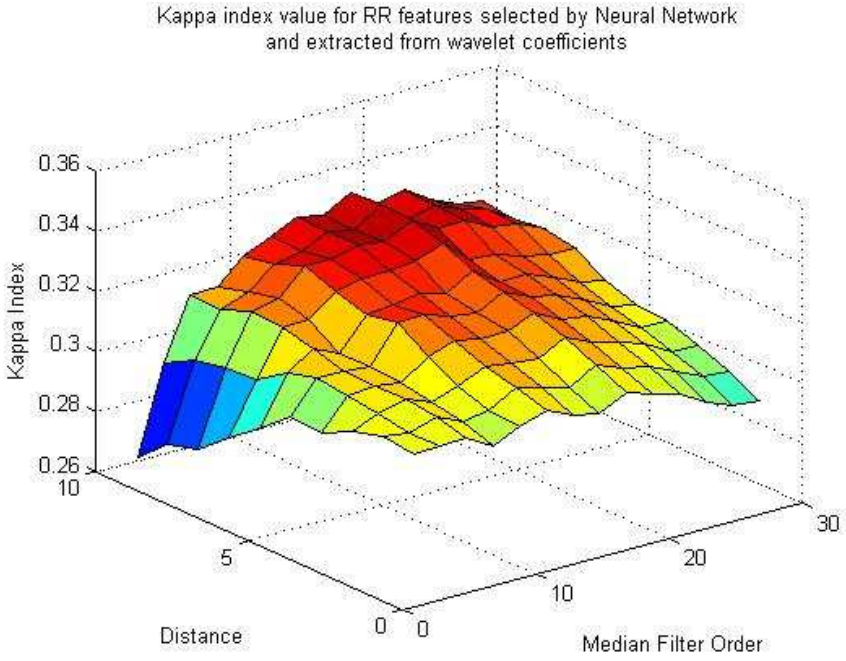


Figure 5.26: Median filter degree and optimal distances based on mean kappa index for each subject computing with FFNN and using RR features extracted with wavelet method. Median filter order=19 and distance=9 maximize mean kappa index value.

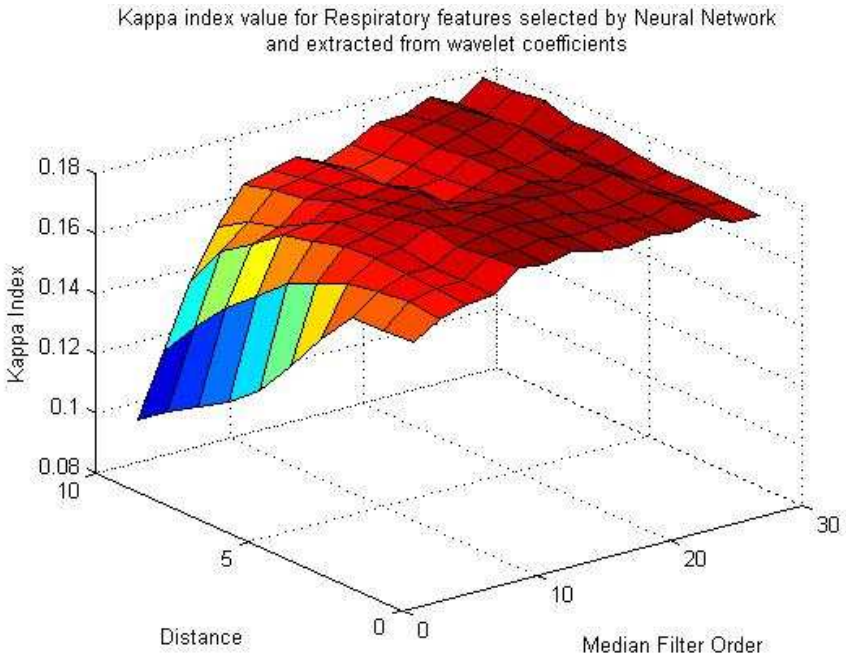


Figure 5.27: Median filter degree and optimal distances based on mean kappa index for each subject computing with FFNN and using respiratory features extracted with wavelet method. Median filter order=29 and distance=8 maximize mean kappa index value.

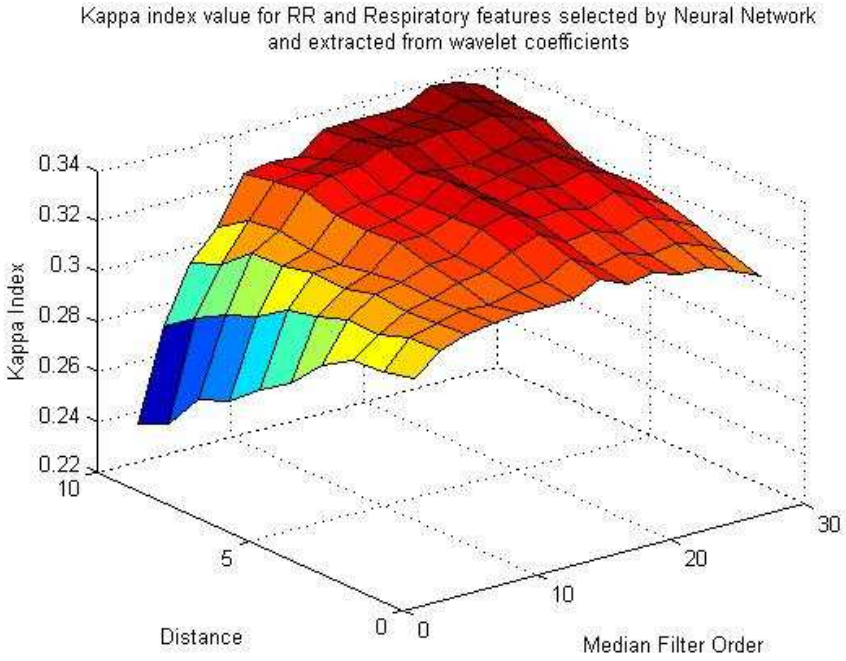


Figure 5.28: Median filter degree and optimal distances based on mean kappa index for each subject computing with FFNN and using RR and respiratory features extracted with wavelet method. Median filter order=25 and distance=10 maximize mean kappa index value.

RLS based features:	WAKE		NREM		REM		TOTAL	
	Accuracy	Kappa index	Accuracy	Kappa index	Accuracy	Kappa index	Accuracy	Kappa index
RR	0,833	0,075	0,689	0,384	0,757	0,222	0,551	0,271
Respiratory	0,831	0,116	0,573	0,260	0,751	0,202	0,444	0,209
RR + Respiratory	0,806	0,147	0,731	0,424	0,793	0,260	0,626	0,325
RR + movement	0,858	0,353	0,694	0,398	0,768	0,237	0,564	0,298
Respiratory + movement	0,853	0,357	0,577	0,268	0,761	0,214	0,455	0,228
RR + Respiratory + movement	0,825	0,338	0,738	0,444	0,802	0,274	0,639	0,356
RR + movement + smooth	0,845	0,037	0,710	0,428	0,799	0,260	0,613	0,357
Respiratory + movement + smooth	0,837	0,043	0,564	0,259	0,787	0,226	0,479	0,255
RR + Respiratory + movement + smooth	0,819	0,129	0,747	0,458	0,826	0,288	0,674	0,406

Table 5.5a: Single state and 3-state accuracy are presented. The last column on the right shows mean kappa index value after merging. The best result is with RR and respiratory RLS based feature combination applying optimal median filter (order=15).

RLS based features:	WAKE vs NREM+REM		NREM vs WAKE+REM		REM vs WAKE+NREM	
	Accuracy	Kappa index	Accuracy	Kappa index	Accuracy	Kappa index
RR	0,643	0,112	0,693	0,395	0,765	0,234
Respiratory	0,560	0,119	0,574	0,262	0,755	0,208
RR + Respiratory	0,720	0,198	0,735	0,436	0,797	0,260
RR + movement	0,655	0,178	0,697	0,407	0,775	0,247
Respiratory + movement	0,570	0,164	0,577	0,270	0,764	0,220
RR + Respiratory + movement	0,732	0,272	0,742	0,455	0,805	0,273
RR + movement + smooth	0,685	0,258	0,732	0,458	0,809	0,274
Respiratory + movement + smooth	0,577	0,205	0,591	0,293	0,789	0,225
RR + Respiratory + movement + smooth	0,753	0,352	0,762	0,489	0,833	0,294

Table 5.5b: 2 states accuracy and kappa index for each class against all after merging.

Features are extracted with RLS method.

RLS based features:	Sleep efficiency		REM during TST		1st REM Latency	
	R&K	Automatic	R&K	Automatic	R&K	Automatic
RR	0,852	0,654	0,234	0,293	79,618	32,029
Respiratory	0,852	0,509	0,234	0,428	79,618	31,765
RR + Respiratory	0,852	0,726	0,234	0,216	79,618	39,853
RR + movement	0,852	0,634	0,234	0,284	79,618	33,353
Respiratory + movement	0,852	0,494	0,234	0,422	79,618	31,147
RR + Respiratory + movement	0,852	0,706	0,234	0,209	79,618	39,941
RR + movement + smooth	0,852	0,647	0,234	0,195	79,618	183,471
Respiratory + movement + smooth	0,852	0,494	0,234	0,354	79,618	147,059
RR + Respiratory + movement + smooth	0,852	0,700	0,234	0,144	79,618	152,882

Table 5.5c: Medical and automatic sleep efficiency comparing.

Wavelet based features:	WAKE		NREM		REM		TOTALE	
	Accuracy	Kappa index	Accuracy	Kappa index	Accuracy	Kappa index	Accuracy	Kappa index
RR	0,845	0,195	0,690	0,349	0,772	0,205	0,580	0,270
Respiratory	0,855	0,173	0,557	0,207	0,763	0,095	0,400	0,140
RR + Respiratory	0,848	0,263	0,671	0,337	0,755	0,222	0,556	0,260
RR + movement	0,859	0,342	0,691	0,355	0,778	0,213	0,586	0,282
Respiratory + movement	0,870	0,352	0,558	0,207	0,767	0,102	0,404	0,147
RR + Respiratory + movement	0,855	0,343	0,672	0,341	0,760	0,228	0,560	0,270
RR + movement + smooth	0,853	0,045	0,714	0,387	0,808	0,173	0,649	0,331
Respiratory + movement + smooth	0,852	-0,001	0,549	0,228	0,805	0,066	0,413	0,180
RR + Respiratory + movement + smooth	0,846	0,018	0,672	0,331	0,818	0,249	0,654	0,339

Table 5.5d: Single state and 3-state accuracy are presented. The last column on the right shows mean kappa index value after merging. The best result is with RR wavelet based feature combination applying optimal median filter (order=25).

Wavelet based features:	WAKE vs NREM+REM		NREM vs WAKE+REM		REM vs WAKE+NREM	
	Accuracy	Kappa index	Accuracy	Kappa index	Accuracy	Kappa index
RR	0,692	0,195	0,694	0,363	0,775	0,205
Respiratory	0,477	0,084	0,558	0,210	0,764	0,098
RR + Respiratory	0,678	0,163	0,675	0,348	0,758	0,225
RR + movement	0,697	0,223	0,696	0,367	0,780	0,213
Respiratory + movement	0,481	0,097	0,559	0,210	0,768	0,104
RR + Respiratory + movement	0,682	0,186	0,676	0,351	0,763	0,232
RR + movement + smooth	0,748	0,329	0,738	0,411	0,812	0,181
Respiratory + movement + smooth	0,433	0,125	0,588	0,279	0,805	0,065
RR + Respiratory + movement + smooth	0,747	0,307	0,739	0,412	0,821	0,249

Table 5.5e: Bi-state accuracy and kappa index, using as classes WAKE against all and NREM against all. Medical sleep efficiency is compared to estimated. Features are extracted from wavelet coefficients.

Wavelet based features:	Sleep efficiency		REM during TST		1st REM Latency	
	R&K	Automatic	R&K	Automatic	R&K	Automatic
RR	0,852	0,677	0,234	0,240	79,618	16,324
Respiratory	0,852	0,418	0,234	0,276	79,618	28,529
RR + Respiratory	0,852	0,676	0,234	0,317	79,618	19,588
RR + movement	0,852	0,668	0,234	0,233	79,618	16,676
Respiratory + movement	0,852	0,413	0,234	0,269	79,618	30,441
RR + Respiratory + movement	0,852	0,667	0,234	0,312	79,618	19,971
RR + movement + smooth	0,852	0,698	0,234	0,106	79,618	159,412
Respiratory + movement + smooth	0,852	0,328	0,234	0,067	79,618	311,676
RR + Respiratory + movement + smooth	0,852	0,723	0,234	0,130	79,618	176,735

Table 5.5f: Medical and automatic sleep efficiency comparing

Results obtained with RLS feature extractor are shown in tables 5.5a/b/c, while results obtained with wavelet decomposition are shown in tables 5.5d/e/f. The best classification is obtained using movement for WAKE stages identification with RR and respiratory features extracted from RLS model (WAKE interval=10 and best median filter order=13) and wavelet coefficients (WAKE interval=10 and median filter order =25), as shown in the previous figures. Figure 5.29 shows an example of automatic sleep staging using RR and Respiratory features extracted from RLS model (middle box) parameters and wavelet coefficients (bottom box) in according to R&K methods (upper box).

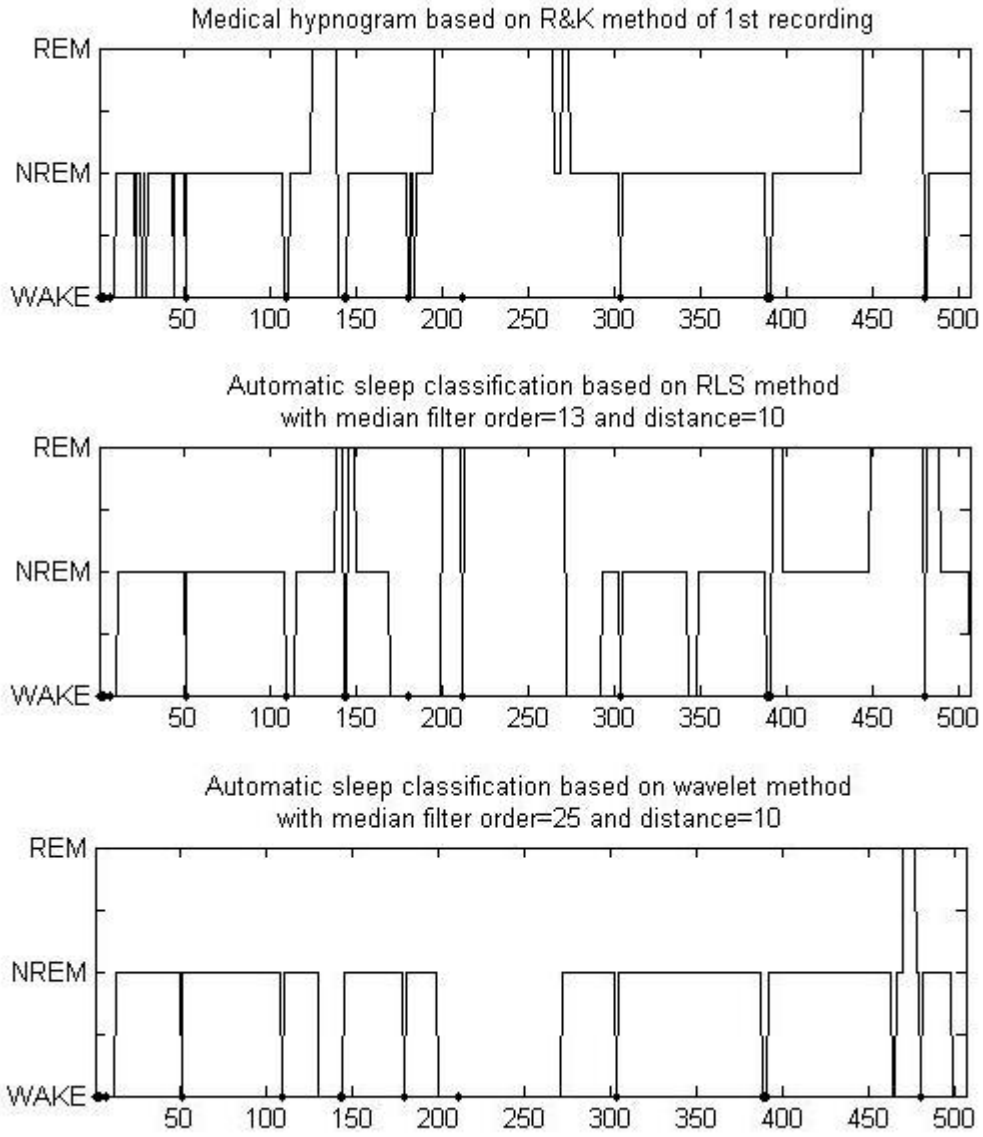


Figure 5.29: The upper plot shows medical hypnogram according to R&K method. In the middle plot, an example of automatic hypnogram using RLS and wavelet (below) based LD feature selection.

Chapter 6

Discussion and Conclusions

"Very little can be said about sleep
that has not been said already"

Nathaniel K.

The aim of this study was the research an automatic algorithm capable of distinguishing the sleep stages from the HRV and respiratory signals. In order to accomplish this task, a detailed analysis of the HRV signal was performed, identifying the most informative features related to the sleep stages. Such features not only allowed us to implement an automatic detection algorithm, but they could lead to a better understanding of the physiological phenomena underlying the sleep process. The algorithm performances were evaluated using the agreement between expert score and the automatic algorithms and the following clinical measures were calculated: sleep efficiency, REM during Total Sleep Time (TST) and the first REM cycle latency.

Several studies have implemented similar algorithms with different techniques in the past:

- Karlen *et al*. [77] implemented a method for the online classification of sleep/wake staes based on cardiorespiratory signals produced by wearable sensors. The method uses a Fast Fourier Transform as the main feature extraction tool and a feed-forward Artificial Neural Network as a classifier. When the method is applied to data collected from a single young male adult, the system can correctly classify on average 95.4% of unseen data from the same user. When the method is applied to classify data from multiple users with the same age and gender, its accuracy is reduced to 85.3%. The proposed method produces a more balanced correct classification of sleep and wake periods than actigraphy. Additionally, by adjusting

the classification threshold of the neural classifier, they obtained 86.7% of correct classification.

- M.O Mendez *et al*. [74] used an algorithm to evaluate the sleep macrostructure based on heart rate fluctuations from ECG signal. The algorithm evaluates the sleep quality out of sleep centers. They used a time-variant autoregressive model as feature extractor and a hidden Markov model as classifier. Characteristics coming from the joint probability of HRV features were used to fed the HMM. Automatic sleep classification is compared to hypnogram given by experts, reaching a total accuracy of 78.21±6.44% and a kappa index of 0.41±.1085 using two features and a total accuracy of 79.43±8.83% and kappa index of 0.42±.1493 using three features.
- Rajeev Agarwal *et al*. [78] presented a computer-assisted sleep staging method. The method uses the principles of segmentation and self-organization (clustering) based on primitive sleep-related features to find the pseudo natural stages present in the record. Sample epochs of these natural stages are presented to the user, who can classify them according to the R&K or any other standard. The method then learns from these samples to complete the classification. Results showed an overall concurrence of 80.6% with manual scoring of 20-s epochs according to R&K standard. The greatest amount of errors occurred in the identification of the highly transitional Stage 1, 54% of which was misclassified into neighboring stages 2 or Wake.
- Redmond *et al*. [17] recently investigated the possibility of obtaining simplified Sleep-Wake-REM sleep stage Information from subjects being assessed for Obstructive Sleep Apnea Syndrome (OSAS), using only electrocardiogram and respiration signals. Their study examined a database of 31 male subjects (Age=42.0±7.4 years) with PSG recording, electrocardiogram signal and an inductance plethysmogram estimate of respiratory. They compare the performances of both linear and quadratic discriminant classifiers in labelling 30-second epochs. The best performance obtained was achieved by a linear discriminant classifier model using a time-dependent *a priori* probability. With a 3-class (W, N, R) system the classification achieved an agreement of kappa index of 0.45 and of 0.57 when a

simplified 2-class (W, S/R) system is considered. This corresponds to an epoch sleep-wake classification accuracy of 89%.

The peculiarity of our study, with respect to those reported above, is that we used signals (HRV, breath and movement) recorded by a smart bed, with approximately six hours of sleep during daytime sleep after a night shift of work and during night time sleep. Our method was supervisioned by the hypnogram provided by the clinicians. Furthermore, two different approaches for feature extraction were used and compared: Autoregressive modeling RLS and wavelet. Our main observation are: a)the combination of HRV and respiratory features extracted with RLS method have the best concordance with medical scoring, b) from the obtained feature set, the best performing feature subset was performed using a Linear Discriminant Classifier and c) Movement activity was used to stage epochs as WAKE, improving classification performances.

The best performance classification obtained for each of these methods, it is reported in table 6.1.

	Feature extracted by RLS method							
	Acc	**kappa**	**SeAuto**	**SeHyp**	**REM%TST auto**	**REM%TST hyp**	**Lat Auto**	**Lat Hyp**
LD	76.39 ± 7.61	0.54 ± 0.10	92.33 ± 2.57	85.21 ± 7.25	29.39 ± 15.50	23.42 ± 6.71	87 ± 49	79 ± 33
QD	76.81 ± 7.51	0.55 ± 0.10	90.25 ± 4.67	85.21 ± 7.25	30.02 ± 14.49	23.42 ± 6.71	86 ± 52	79 ± 33
K-NN	71.95 ± 7.47	0.42 ± 0.10	78.77 ± 10.06	85.21 ± 7.25	6.80 ± 8.02	23.42 ± 6.71	201±149	79 ± 33
FFNN	67.17 ± 11.88	0.39 ± 0.13	71.60 ± 17.62	85.21 ± 7.25	11.60 ± 14.14	23.42 ± 6.71	200±141	79 ± 33

	Feature extracted by Discrete Wavelet Transform							
	Acc	**kappa**	**SeAuto**	**SeHyp**	**REM%TST auto**	**REM%TST hyp**	**Lat Auto**	**Lat Hyp**
LD	73.92 ± 15.33	0.51 ± 0.17	92.27 ± 2.48	85.21 ± 7.25	31.31 ± 22.25	23.42 ± 6.71	78±58	79 ± 33
QD	69.27 ± 12.64	0.35 ± 0.15	91.65 ± 3.25	85.21 ± 7.25	19.36 ± 25.54	23.42 ± 6.71	104±85	79 ± 33
K-NN	64.75 ± 9.30	0.33 ± 0.10	64.55 ± 14.63	85.21 ± 7.25	0.75 ± 1.10	23.42 ± 6.71	239±122	79 ± 33
FFNN	50.37 ± 18.79	0.24 ± 0.13	48.75 ± 23.47	85.21 ± 7.25	22.08 ± 28.31	23.42 ± 6.71	87±122	79 ± 33

Table 6.1. Mean and standard deviation of accuracy and agreement measure for the sleep staging obtained by the four classifiers with the two feature extraction methods. Acc means general accuracy, kappa is kappa index, Se is the sleep efficiency, REM%TST is the percentage of Total Sleep Time (TST) occupied by REM stages, and Lat represents the latency, in minutes, of the first REM cycle. "Auto" suffixed to one of the previous abbreviations means that the parameter is obtained by the automatic system, despite the suffix "hyp" relates parameters obtained from standard hypnogram.

LD and QD are optimized with sequential forward feature selection. QD achieved a better performance with features extracted from RLS model, while it decreased with wavelet based feature extraction. LD maintained same performance for both feature extractor methods used. One important difference between these methods is the number of features, while QD has used 15 features derived from HRV and respiratory signals, LD has used only 8 features, which implies advantage from the computational cost.

The results show that Feed-Forward Neural Network and *K*-Nearest-Neighbor not reached good performances probably due to optimization on linear discrimant, features selection and choice of initial stage number as WAKE.

During this work a SVM classifier was trained on the training set with a search grid technique to optimize parameters. The high computational complexity and the common kernel (polynomial and Gaussian) did not provide good results. The next step forward for this research would be implementation of precomputed kernel on training data.

Both the linear discriminant and quadratic discriminant show good mean kappa index values on each subject and recognition of first REM latency, while the Neural Network leads to a better REM recognition but to a slightly lower mean kappa index value. The neural network architecture used in this study was formed by one hidden layer and two neurons in the output layer. *K*-Nearest-Neighbor leads to good accuracy but to a lower kappa index. It could be interesting to change neural architecture, for example adding more hidden layers, and to use neural network or *K*-NN as classifier in features selecting.

Chapter 7

Bibliography

[1] **Michael B Russo, Shehnaz Shaikh**. *Sleep: Understanding the Basics* pp.1 [http://www.emedicinehealth.com]

[2] **A. M. Bianchi, O. P. Villantieri**. Studio di “vital signs” utili all’identificazione dell’addormentamento alla guida. *Progetto NIDO2*. Dipartimento di Bioingegneria del Politecnico di Milano. 2005

[3] Nervous System. *Columbia Encyclopedia*. Columbia University Press.

[4] **Regina Bailey**. Peripheral Nervous System. *About.com Guide*

[5] **Minors, Waterhouse**. Circadian rhythms in general. *Occup Med* pag:165-182. 1990 Apr-Jun.

[6] **A. C. Guyton, J.E. Hall**. Fisiologia medica. Ed. *EdiSES*. 2002. Trad. it. 9a ed. Napoli, 2003.

[7] **Rechtschaffen, A. and Kales, A.** A manual of standardized terminology, tecniques and scoring system for sleep stages in human subject. *ser. National Instututes of Health Publications.* US Government Printing Office, 1968, p. 204.

[8] **Kushida CA, Littner MR, Morgenthaler TM, et al.**. Practice parameters for the indications for polysomnography and related procedures: An update for 2005. *Sleep 28*: 499-519.

[9] **S. Cerutti U.J. Scholz, A.M. Bianchi and S. Kubicki**. Monitoring of the vegetative background of sleep by means of spectral analysis of heart rate variability. *Computers in Cardiology* 1993. Proceedings., pages 321# 324, 1993.

[10] **A. Malliani, M. Pagani, F. Lombardi, and S. Cerutti**. Cardiovascular neural regulation explored in the frequency domain. *Circulation*, 84(2):482#492, 1991.

[11] **Armour, J. A. and J. Ardell, Eds**. (1994). Neurocardiology. New York, NY, Oxford

University Press.

[12] **M.V. Kamath and E.L. Fallen**. Power spectral analysis of heart rate variability: a noninvasive signature of cardiac autonomic function. *Crit Rev Biomed Eng*, 21(3): 245#311, 1993.

[13] **JC Principe, SK Gala, and TG Chang**. Sleep staging automaton based on the theory of evidence. *Biomedical Engineering, IEEE Transactions on*, 36(5):503#509, 1989.

[14] **F. Poree, A. Kachenoura, H. Gauvrit, C. Morvan, G. Carrault, and L. Senhadji**. Blind Source Separation for Ambulatory Sleep Recording. *IEEE Trans Inf Technol Biomed*, 10(2):293#301, 2006.

[15] **A. Flexer, G. Gruber, and G. Dorfner**. A reliable probabilistic sleep stager based on a single EEG signal. *Artifcial Intelligence In Medicine*, 33(3):199#207, 2005.

[16] **R. Agarwal, J. Gotman, S. Syst, and Q. Montreal**. Computer-assisted sleep staging. *Biomedical Engineering, IEEE Transactions on*, 48(12):1412#1423, 2001.

[17] **S.J. Redmond, P. de Chazal, C. O'Brien, S. Ryan, W.T. McNicholas, and C. Heneghan**. Sleep staging using cardiorespiratory signals. *Somnologie-Schlaforschung und Schlafmedizin*, 11(4):245#256, 2007.

[18] **T. Watanabe and K. Watanabe**. Noncontact method for sleep stage estimation. *Biomedical Engineering, IEEE Transactions on*, 51(10):1735#1748, 2004.

[19] **F. Suozzo, A. e F.Suozzo**. Gli HMM come approccio alla classificazione degli stadi di sonno sulla base del segnale di variabilità cardiaca. Politecnico di Milano, 2006.

[20] **A. Lewicke, E. Sazonov, MJ Corwin, M. Neuman, and S. Schuckers**. Sleep Versus Wake Classification From Heart Rate Variability Using Computational Intelligence: Consideration of Rejection in Classification Models. *Biomedical Engineering, IEEE Transactions on*, 55(1):108#118, 2008.

[21] **Dennis Ward Ricker** (2003). Echo Signal Processing. *Springer*. ISBN140207395X.

[22] **Priestley, M.B**. (1991). Spectral Analysis and Time Series. *Academic Press*. ISBN 0-12-564922-3.

[23] **A.M. Bianchi et al**. Batch and Time-variant parametric spectral analysis for respiratory component estimation in heart rate variability signal. *J. Ambulatory Monitoring*, 5:107#21, 1992.

[24] **O. Rioul, V. Martin**. Wavelets and Signal Processing. *IEEE SP Magazine*, pp. 14-38, October 1991.

[25] **M. G . Amin**. Timefrequency spectrum analysis and estimation for non-stationary random processes. In Boualem Boashash, editor. *Time- Frequency Signal Analysis*, chapter 9, pages 208- 232. Longman Cheshire, 1992.

[26] **B. Boashash**. Time-frequency signal analysis. In S. Haykin, editor, Advances in Spectral Estimation and Array Processing, volume 1 of 2, chapter 9, pages 418-517. Prentice Hall, Englewood Cliffs, New Jersey, 1990.

[27] **M. Roessgen, M. Deriche, and B. Boashash**. A Comparative Study of Spectral Estimation Techniques for Noisy Non-Stationary Signals with Application to EEG Data . Signal Processing Research Centre Queensland University of Technology Brisbane, Q., 4001, Australia . ISBN 1058-6393/9

[28] **A.M. Bianchi**. MyHeart TakeCare: Algorithms for the Sleep Module. Politecnico di Milano, 2006

[29] **L. Guo, L. Ljung and T. Priouret**. Performance analysis of the forgetting factor RLS algorithm. *International journal of adaptive control and signal processing*, 7(6):525537, 1993.

[30] **A. Graps** .An Introduction to Wavelets,. *IEEE Computational Sciences and Engineering* 2001.

[31] **Mallat S**. A wavelet tour of signal processing. 2nd ed. San Diego: Academic Press; 1999.

[32**] J. Ooi and V. Viswanathan**. Applications of Wavelets to Speech Processing. Modern Methods of Speech Processing, R.P. Ramachandran and R. Mammone, ed., Kluwer Academic Publishers, Boston, 1995, pp. 449-464.

[33] **V. Viswanathan, W. Anderson, J. Rowlands, M. Ali and A. Tewfik**. Real-Time Implementation of a Wavelet-Based Audio Coder on the T1 TMS320C31 DSP Chip,. 5th International Conference on Signal Processing Applications & Technology (ICSPAT), Dallas, TX, Oct. 1994.

[34] **M. Misiti, Y. Misiti, G. Oppenheim and J. Poggi.** Matlab Wavelet Tool Box, The Math Works Inc., 2009.

[35] **Altman D.G**. Practical Statistcis for Medical Research, Chapman & Hall, 1991

[36] **R. A. Fisher**. Statistical Methods for Research Workers, Edinburgh: Oliver and Boyd, 1925, p.43.

[37] **Kruskal and Wallis**. Use of ranks in one-criterion variance analysis. *Journal of the*

American Statistical Association 47 (260): 583–621, December 1952.

[38] **Bray, J. H., & Maxwell, S. E.** Multivariate analysis of variance. *Newbury Park*, CA: Sage Publications.

[39] The log transformation is special. Keene ON. Stat Med 1995: 14(8); 811-9.

[40] http://en.wikipedia.org/wiki/Lognormal_distribution

[41] **David Sheskin**. Handbook of parametric and nonparametric statistical procedures: 405-406

[42] **G.G. Berntson, J.T. Bigger Jr., D.L. Eckberg, P. Grossman, P.G. Kaufmann, M. Malik, H.N. Nagaraja, S.W. Porges, J.P. Saul, P.H. Stone, and M.W. Van Der Molen**. Heart rate variability: Origins, methods, and interpretive caveats. Psychophysiol, 34:623–648,1997.

[43] **A. Malliani, M. Pagani, F. Lombardi, and S. Cerutti**. Cardiovascular neural regulation explored in the frequency domain. *Circulation*, 84(2):482–492, August 1991.

[44] Task force of the European society of cardiology and the North American society of pacing and electrophysiology. Heart rate variability – standards of measurement, physiological interpretation, and clinical use. Circulation, 93(5):1043–1065, March 1996.

[45] **H.-J. Braune and U. Geisen¨orfer**. Measurement of heart rate variations: influencing factors, normal values and diagnostic impact on diabetic autonomic neuropathy. Diabetes Res Clin Practice, 29:179–187, 1995.

[46] **M. Pagani**. Heart rate variability and autonomic diabetic neuropathy. Diabetes Nutrition & Metabolism, 13(6):341–346, 2000.

[47] **F. Lombardi, T.H. M¨akikallio, R.J. Myerburg, and H. Huikuri**. Sudden cardiac death: role of heart rate variability to identify patients at risk. Cardiovasc Res, 50:210–217, 2001.

[48] **H.V. Huikuri, T.H. M¨akikallio, P. Raatikainen, J. Perki¨om¨aki, A. Castellanos, and R.J. Myerburg**. Prediction of sudden cardiac death: appraisal of the studies and methods assessing the risk of sudden arrhythmic death. *Circulation*, 108(1):110–115, July 2003.

[49] **O. Rompelman**. Rhythms and analysis techniques. In J. Strackee and N. Westerhof, editors, *The Physics of Heart and Circulation*, pages 101–120. Institute of Physics Publishing, Bristol, 1993.

[50] **N.V. Thakor, J.G. Webster, and W.J. Tompkins**. Optimal QRS detector. *Med Biol*

Eng Comput, 21:343–350, May 1983.

[51] **O. Pahlm and L. S¨ornmo**. Software QRS detection in ambulatory monitoring – a review. *Med Biol Eng Comput*, 22:289–297, July 1984.

[52] **J. Pan and W.J. Tompkins**. A real-time QRS detection algorithm. *IEEE Trans Biomed Eng*, 32(3):230–236, March 1985.

[53] **P.S. Hamilton and W.J. Tompkins**. Quantitative investigation of QRS detection rules using the MIT/BIH arrhythmia database. *IEEE Trans Biomed Eng*, 33(12):1157–1165, December 1986.

[54] **G.M. Friesen, T.C. Jannett, M.A. Jadallah, S.L. Yates, S.R. Quint, and H.T. Nagle**. A comparison of the noise sensitivity of nine QRS detection algorithms. *IEEE Trans Biomed Eng*, 37(1):85–98, January 1990.

[55] **V.X. Afonso**. ECG QRS detection. In W.J. Tompkins, editor, *Biomedical Digital Signal Processing*, chapter 12, pages 237–264. Prentice Hall, New Jersey, 1993.

[56] **M. Merri, D.C. Farden, J.G. Mottley, and E.L. Titlebaum**. Sampling frequency of the electrocardiogram for spectral analysis of the heart rate variability. *IEEE Trans Biomed Eng*, 37(1):99–106, January 1990.

[57] **G.D. Pinna, R. Maestri, A. Di Cesare, R. Colombo, and G. Minuco**. The accuracy of power-spectrum analysis of heart-rate variability from annotated RR lists generated by Holter systems. *Physiol Meas*, 15:163–179, 1994.

[58] **I. Daskalov and I. Christov**. Improvement of resolution in measurement of electrocardiogram RR intervals by interpolation. *Med Eng Phys*, 19(4):375–379, June 1997.

[59] **T. Bragge, M.P. Tarvainen, P.O. Ranta-aho, and P.A. Karjalainen**. High-resolution QRS fiducial point corrections in sparsely sampled ECG recordings. *Physiol Meas*, 26(5):743–751, 2005.

[60] **O. Rompelman**. Rhythms and analysis techniques. In J. Strackee and N. Westerhof, editors, *The Physics of Heart and Circulation*, pages 101–120. Institute of Physics Publishing, Bristol, 1993.

[61] **G. Baselli, S. Cerutti, S. Civardi, F. Lombardi, A. Malliani, M. Merri, M. Pagani, and G. Rizzo**. Heart rate variability signal processing: a quantitative approach as an aid to diagnosis in cardiovascular pathologies. *Int J Bio-Med Comput*, 20:51–70, 1987.

[62] **J. Mateo and P. Laguna**. Improved heart rate variability signal analysis from the beat

occurrence times according to the IPFM model. *IEEE Trans Biomed Eng*, 47(8):985–996, August 2000.

[63] **R.W. DeBoer, J.M. Karemaker, and J. Strackee**. Comparing spectra of a series of point events particularly for heart rate variability data. *IEEE Trans Biomed Eng*, 31(4):384–387, April 1984.

[64] **R.W. DeBoer, J.M. Karemaker, and J. Strackee**. Spectrum of a series of point events, generated by the integral pulse frequency modulation model. *Med Biol Eng Comput*, 23:138–142, March 1985.

[65**] N. Lippman, K.M. Stein, and B.B. Lerman**. Nonlinear predictive interpolation: a new method for the correction of ectopic beats for heart rate variability analysis. J *Electrocardiol*, 26:S14–S19, 1993.

[66] **N. Lippman, K.M. Stein, and B.B. Lerman**. Comparison of methods for removal of ectopy in measurement of heart rate variability. *Am J Physiol*, 267(1):H411–H418, July 1994.

[67**] J. Mateo and P. Laguna**. Analysis of heart rate variability in the presence of ectopic beats using the heart timing signal. *IEEE Trans Biomed Eng*, 50(3):334–343, March 2003.

[68] **J. Malmivuo and R. Plonsey**. Bioelectromagnetism: Principles and Applications of Bioelectric and Biomagnetic Fields. *Oxford University Press* (Web Edition), 1995.

[69] **Widrow, B. and M. A. Lehr**. 30 years of adaptive neural networks: perceptron, madeline, and backpropagation. *Proceedings of the IEEE* 78(9), 1415–1442.

[70] **McCulloch, W. S. and W. Pitts**. A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics* 5, $115–133. Reprinted in Anderson and Rosenfeld (1988).

[71] **Rosenblatt, F**.. Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms. *Spartan*.

[72] **Hagan, M.T., and M. Menhaj.** Training feed-forward networks with the Marquardt algorithm. *IEEE Transactions on Neural Networks*, Vol. 5, No. 6, 1999, pp. 989–993, 1994.

[73] **H. Demuth, M. Beale and Martin Hagan.** Matlab Neural Network Tool Box, The Math Works Inc., 2009.

[74] **Juha M. Kortelainen, Martin O. Mendez, Anna M. Bianchi, Matteo Matteucci and Sergio Cerutti**. Staging based on Signals Acquired through Bed Sensor. *IEEE Trans Biomed Eng*

[75] **R. O. Duda, P. E. Hart, D. G. Stork.** Pattern Classification. *Wiley-Interscience*. 2000.

[76] **C. M. Bishop**. Pattern Recognition and Machine Learning. 2006.

[77] **Karlen, W, Mattiussi, C and Floreano, D** (2009) Sleep and Wake Classification With ECG and Respiratory Effort Signals. *IEEE Transactions on Biomedical Circuits and Systems.*

[78] **Rajeev Agarwal and Jean Gotman**. Computer-Assisted Sleep Staging. *IEEE Transactions on Biomedical Engineering*, Vol. 48

[77] **Karlen W., Mattiussi C., and Floreano D.**. Sleep and Wake Classification With ECG and Respiratory Effort Signals. *IEEE Transactions on Biomedical Circuits and Systems*.

[78] **D. P. White.** Sleep Apnea. *Proc. Am. thorac. Soc*., vol. 3, pp. 124-128, 2006.

[79] **T. Young, P. E. Peppard and D. G. Gottlier.** Epidemiology of obstructive sleep apnea, a population health perspective. *Am. J. Respir. Crit. Care. Med*., vol. 165, pp. 1217-1239, 2002.

[80] Report of American Academy of sleep Medicine Task Force. Sleep-related breathing disorders in adults: recommendations for syndrome definition and measurement techniques in clinical research. *Sleep*, vol. 22 no. 5, pp. 667-689, Aug. 1999.

[81] **J. Kortelainen, J. Virkkala**. FFT averaging of multichannel BCG signals from bed mattress sensor to improve estimation of heart beat interval., *29th IEEE EMBS Annual International Conference*, August, pp. 23-26, 2007, Lyon, France.

[82] **M. Nitzan, A. Babchenko, B. Khanokh, and D. Landau.** The variability of the photoplethysmographic signal-a potential method for the evaluation of the autonomic nervous system, *Physiol. Meas*., vol. 19, pp. 93102, 1998.

[83] **J. Allen.** Photoplethysmography and its Application in Clinical Physiological Measurement. *Physiol. Meas*., vol. 28, pp. 1-39, 2007.

[84] **R. Paradiso, G. Loriga, N. Taccini.** A Wearable Health Care System Based on Knitted Integrated Sensors, *Information Technology in Biomedicine, IEEE Transactions on*, vol. 9, pp. 337344, 2005.

[85] **M. H. Bonnet and D. L. Arand.** Heart rate variability: sleep stage, time of night, and arousal influences. *Electroencephalogr. Clin. Neurophysiol*., vol. 102, no. 5, pp. 390-396, May. 1997.

[86] **M. Brink, C.H. Mueller and C. Schierz** Contact-free Measurement of Heart Rate, Respiration Rate, and Body Movements during Sleep. *Behavior Research Methods*, vol. 38, pp. 511-521, 2006.

[87] **J. Alihanka, K. Vaahtoranta and I. Saarikivi.** A new method for long-term monitoring of the ballistocardiogram, heart rate, and respiration. *Am J Physiol Regulatory Integrative Comp Physiol* 240:384-392, 1981.