Automatic Detection of Sleep Transients and Its Applications in Sleep Spindle Enhancement

A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

Chanakya Reddy Patti
MEng – RMIT University, BEng – Monash University

School of Engineering
College of Science, Engineering and Health
RMIT University

June 2018
Declaration

I certify that except where due acknowledgement has been made, the work is that of the author alone; the work has not been submitted previously, in whole or in part, to qualify for any other academic award; the content of the thesis is the result of work which has been carried out since the official commencement date of the approved research program; any editorial work, paid or unpaid, carried out by a third party is acknowledged; and, ethics procedures and guidelines have been followed.

I acknowledge the support I have received for my research through the provision of an Australian Government Research Training Program Scholarship.

Chanakya Reddy Patti

23-05-2018
Acknowledgments

This thesis would not have been possible without the support of a many individuals and institutions. I take this opportunity to thank these individuals and institutions who have in one way or another made a contribution.

First and foremost, my sincerest thanks to my PhD adviser, Dr Dean Cvetkovic, without his continuous inputs and support this thesis would not have been possible.

I am indebted to my PhD co-advisor, Professor Hong Ren Wu, whose inputs at very critical junctures of my PhD journey have helped me find the necessary path and motivation to keep going. As to how someone of such few words has come to highly influence my views on research and life is something that has been a puzzle to me.

I would like to express my gratitude to Dr Elena Pirogova for her external guidance and support during my PhD.

My sincere thanks to Professor Thomas Penzel from the Charité Universitätsmedizin Berlin for collaboration and providing me with an opportunity to conduct part of my PhD research in Germany.

I would like to acknowledge and thank the following institutions or scholarship programs which have provided financial support in the form of scholarships during my PhD:

- Australian Government Research Training Program Scholarship.
- RMIT University
- German Academic Exchange Service (DAAD) Scholarship.

And finally, I would like to acknowledge the support, emotional or other, I have received from family and friends which have helped me focus on my PhD while excluding external problems. I take this opportunity to individually acknowledge the support of my parents, Dr. Narasimha Reddy Patti and Swarnalatha Mula, family and more- Janardhana Reddy and Vijaya Lakshmi; Kumar Deshmukh and Sunanda Deshmukh; Sreenivas Reddy; Dr. Bhadra Reddy and Dr. Srilatha Reddy; Dr. John Pragastis and Chrysante Pragastis.
Contents
Declaration............................................................................................................................................i
Acknowledgements............................................................................................................................iii
List of Figures.........................................................................................................................................ix
List of Tables..........................................................................................................................................xii
Abstract................................................................................................................................................2
Abbreviations.........................................................................................................................................4
Symbols..................................................................................................................................................7
List of Publications by the Author........................................................................................................10
1  Introduction.........................................................................................................................................12
   1.1  Brief history of sleep research ......................................................................................................12
   1.2  Electroencephalography (EEG) Rhythms....................................................................................13
   1.3  Sleep Stages and Sleep Transients...............................................................................................13
       1.3.1  Sleep Stages.........................................................................................................................13
       1.3.2  Sleep Transients..................................................................................................................15
   1.4  Research Aims..............................................................................................................................17
       1.4.1  Research Objectives.............................................................................................................17
       1.4.2  Hypotheses............................................................................................................................18
       1.4.3  Significance & Limitations:....................................................................................................18
   1.5  Contributions of Thesis Work......................................................................................................19
   1.6  Thesis Outline..............................................................................................................................20
2  Sleep Spindle and K-complex definitions, detection methods, generation mechanisms and significance.................................................................................................................................21
   2.1  Sleep Spindle and K-complex Definitions....................................................................................21
       2.1.1  American Academy of Sleep Medicine definitions............................................................21
       2.1.2  Definitions by Rechtschaffen & Kales..................................................................................22
   2.2  Automatic sleep spindle and K-complex detection methods.......................................................22
       2.2.1  Automatic sleep spindle detection .......................................................................................22
       2.2.2  Automatic K-complex detection.............................................................................................29
   2.3  Sleep spindle generation mechanisms.........................................................................................33
   2.4  Functional significance of sleep spindles.....................................................................................36
       2.4.1  Sleep spindles, hippocampus ripples and memory consolidation........................................36
2.5 Clinical significance of sleep spindles in memory consolidation ........................................37
2.6 Sleep spindles and K-complexes in abnormal conditions ....................................................37
2.7 Sleep spindle and K-complex enhancement methods .............................................................38
  2.7.1 Auditory stimulation to enhance sleep spindles and K-complexes ......................................38
  2.7.2 Neurofeedback to enhance sleep spindles ...........................................................................40
3 Automated sleep transient detection—Fundamentals ...............................................................43
  3.1 Performance Measures ........................................................................................................43
  3.2 Automated sleep spindle detection .......................................................................................44
      3.2.1 Databases ................................................................................................................44
      3.2.2 Clustering technique ..................................................................................................45
      3.2.3 Random Forest Classifier .........................................................................................55
  3.3 Automated K-complex detection ..........................................................................................59
      3.3.1 Databases ................................................................................................................59
      3.3.2 Clustering method .....................................................................................................59
      3.3.3 Detection using pattern matched wavelets .................................................................60
4 Automated sleep transient detection techniques ......................................................................65
  4.1 Sleep spindle detection using STFT .....................................................................................65
      4.1.1 Dreams Database .......................................................................................................65
      4.1.2 MASS database ........................................................................................................68
      4.1.3 Charité Database .....................................................................................................70
  4.2 Sleep spindle detection using IIR filters ...............................................................................72
      4.2.1 Dreams Database .......................................................................................................75
      4.2.2 MASS database ........................................................................................................77
      4.2.3 Charité Database .....................................................................................................85
      4.2.4 Results using the random forest classifier— ...............................................................86
  4.3 K complex detection using clustering method ......................................................................87
      4.3.1 Dreams K-complex database .....................................................................................87
  4.4 K complex detection using Pattern matched wavelet ...........................................................89
      4.4.1 Dreams K-complex database .....................................................................................89
  4.5 Summary Tables of Sleep Transient Detection ...................................................................91
5 Auditory Stimulation and Neurofeedback for Enhancing Sleep Spindle Count .....................92
  5.1 Study Design .......................................................................................................................92
5.1.1 Participants: ................................................................. 92
5.1.2 Methodology: ............................................................ 92
5.2 Materials Design and Apparatus ..................................... 93
  5.2.1 Declarative Memory Task: Paired Associate Learning .... 93
  5.2.2 Auditory stimulation protocol .................................... 94
  5.2.3 EEG recordings ....................................................... 98
  5.2.4 Neurofeedback Protocol ........................................... 98
  5.2.5 Auditory stimulation + neurofeedback protocol .......... 100
  5.2.6 Electroencephalography analyses .............................. 100
6 Auditory Stimulation and Neurofeedback ........................... 101
  6.1 Comparison of first three sessions ............................... 106
7 Discussions and Conclusions ........................................... 108
  7.1 Automatic sleep transient detection .............................. 108
    7.1.1 Sleep spindle detection ....................................... 108
    7.1.2 K-complex detection ........................................... 112
  7.2 Auditory stimulation and neurofeedback protocol .......... 113
    7.2.1 Comparison of the three groups ............................. 113
  7.3 Conclusion .................................................................. 116
Appendix A ...................................................................... 118
Appendix B ...................................................................... 126
Appendix C ...................................................................... 131
Appendix D ...................................................................... 134
References ...................................................................... 144
List of Figures

Figure 1-1 Three fields of studies which underpin the work presented in this dissertation................. 12
Figure 1-2 Transient EEG events seen in wake, NREM-1 and NREM-2. Adapted from [10]............ 14
Figure 1-3 Typical pattern of EOG, EEG and EMG signals observed during REM sleep............ 15
Figure 2-1 Typical K-complex showing a well delineated negative sharp wave immediately followed by a positive component. Maximal amplitude of a K-complex is usually observed using the frontal derivations, i.e. the signals from electrodes placed towards the front of the scalp............................... 20
Figure 2-2 Illustration of a thresholding detection method. An EEG signal (top plot) is filtered in the SS frequency band; this is followed by identifying filtered signal segments whose power is greater than a set threshold (bottom plot)............................................................................................................. 22
Figure 3-1 Signal showing the four STFT features along with the original signal (top plot). A spindle is detected in the region marked ‘Spindle region’.......................................................................................... 47
Figure 3-2 EEG electrode positions in the 10-20 system, adapted from [99].............................. 49
Figure 3-3 Hypothetical decision tree based on three features X,Y,Z used to classify into two classes named abnormal and normal.................................................................................................................. 55
Figure 3-4 Pattern matched wavelet developed using Eq. 45. The process of matching is described in 3.3.3.1 along with the original K-complex used for matching......................................................... 61
Figure 3-5, Pattern matched wavelet #2 developed using Eq. 45.................................................. 61
Figure 4-1 ROC curves for different moving window sizes (Dreams database), the intersection of the curves with the dotted line show the approximate ideal trade-off between Sensitivity and False Positive Proportion. The legend describes the moving window length used for each curve, along with the area under the curve (AUC) for each curve. Curves with higher AUC lean towards the ideal results towards the top left hand corner of the plot................................................................. 65
Figure 4-2 Single ROC curve obtained by changing moving window length (Dreams Database), while keeping the minimum length check to remove false spindles constant at 100% of moving window length .................................................................................................................. 66
Figure 4-3 ROC curves for different moving window sizes (MASS database), the intersection of the curves with the dotted line show the approximate ideal trade-off between Sensitivity and False Positive Proportion.......................................................................................................................... 68
Figure 4-4 ROC curves for different moving window sizes (Charité Database), the intersection of the curves with the dotted line show the approximate ideal trade-off between Sensitivity and False Positive Proportion.......................................................................................................................... 71
Figure 4-5 Comparison of processing time required by the STFT based clustering method and the IIR filters based clustering method, executed on a MATLAB 2015b platform running on an Intel i7-core CPU @ 3.6 Ghz. Processing time for event detection was significantly reduced using IIR filters.... 72
Figure 4-6 Comparison of ROC curves of the STFT and IIR filter methods. ROC curves were obtained by varying length checks to remove false spindles using a 1 second moving window. The IIR filters method shows better performance based on higher AUC......................................................... 73
Figure 4-7 Comparison of ROC curves of the STFT and IIR filter methods (Dreams database). ROC curves were obtained by varying length checks to remove false spindles using a 1 second moving window. In this case, the STFT method showed better performance in terms of higher AUC. ..........75

Figure 4-8 Sensitivity, FPP, precision and F1-score for different montages of the Montreal Archive of Sleep Studies database (MASS) shown in Table 4-7. Highest sensitivity is seen using the C3 channel which also showed the lowest FPP, highest F1-score and highest precision. Poorest results are seen using the m4 montage which includes central and parietal derivations.........................................78

Figure 4-9 Slow and Fast spindle counts for different montages. A higher number of slow spindles can be seen when using the frontal derivations in m2 and a higher number of fast spindles are seen in the parietal derivations; earlier studies examining the slow fast dichotomy showed similar differences between frontal slow spindles and parietal slow spindles [15][16]. ........................................................................79

Figure 4-10 Central frequency distribution of sleep spindles identified by automatic detection and visual scoring. No visually discernible slow-fast dichotomy can be seen in the plots including visual scoring. Slow spindles can be seen in the plots showing automatic detections using different montages, however very few slow spindles can be seen in the plot showing frequency distribution of the visually scored spindles, suggesting that most visual scored spindles were centro-parietal fast spindles. .................................................................80

Figure 4-11 Comparison of ROC curves of the STFT and IIR filter methods (Charite database). ROC curves were obtained by varying minimum length checks (sub-section 3.2.2.4) to remove false spindles using a 1 second moving window.........................................................................................85

Figure 4-12 F1-score results obtained using different window sizes on the Dreams K-complex database.................................................................................................................................87

Figure 4-13 Sensitivity (TPR) vs Precision (PPV) for different detection thresholds on CWT coefficients. Points representing threshold of 400 are marked in the figure. Two lines show different K-complexes chosen for pattern matching i.e matched wavelet #1, matched wavelet #2........................................89

Figure 5-1 Filter to generate pink noise powers ........................................................................................................92

Figure 5-2 The magnitude and phase response of the filter, as can be seen in the magnitude response, the slope in the 20 – 20000 Hz range is approximately 10db/decade, or 3db/octave..................................93

Figure 5-3 Frequency response of the 10.5-15 Hz filter .................................................................................................94

Figure 5-4 Frequency response of the 4-10 Hz filter .....................................................................................................95

Figure 5-5 Frequency response of the 20-40 Hz filter .....................................................................................................95

Figure 5-6 Overview of the auditory stimulation setup. ..................................................................................................96

Figure 5-7 SMR feedback intra-session blocks, a 20 second break between the 170 second periods is presented using a white cross on a black background.................................................................97

Figure 5-8 SMR feedback screen, the green bar represents SMR activity, while the red and blue bars represent artifact activity which the subject is instructed to try and keep low........................................97

Figure 6-1 Change in spindle density from baseline for all groups. The auditory stimulation groups (1 and 2) show a positive change from baseline, whereas the NF only group (3) shows a negative change of baseline. This is contradictory to some findings by other researchers who observed an increase; this is discussed further in Chapter 7........................................................................................................99

Figure 6-2 Changes in the memory consolidation parameter compared to baseline for all groups. Only Group 1 showed a positive change, whereas both the NF groups (2 and 3) show a negative change.100
Figure 6-3 Box plots showing changes in tiredness ratings were compared with the baseline. No significant difference was found between any of the groups. Data taken from Tables 6-7.............103
Figure 6-4 Total spindle count of the first three sessions for all subjects. Acronym definitions, A – Auditory Stimulation, NF – Neurofeedback.................................................................104
List of Tables

Table 2-1 Overview of major research works in automated sleep spindle detection..........................27
Table 4-1 Ideal results obtained using a 1.2 second window, subjects are numbered S1, S2, …, S6. The Dreams database with two visual scorings is limited to six subjects on which a few existing algorithms have tested their performance. .................................................................................................................68
Table 4-2 Ideal results obtained using a 1 second window for the MASS database. Subjects are numbered S1, S2, …, S15.................................................................................................................................................................70
Table 4-3 Ideal results obtained using a 1 second window for the Charité database, subjects are numbered S1, S2, …, S6.............................................................................................................................................................................71
Table 4-4 Comparison between the STFT method and the IIR filters method using the $F_1$-score statistic. The IIR method showed better performance in terms of producing a higher $F_1$-statistic......75
Table 4-5 IIR filters method results obtained using a 1 second moving window, subjects are numbered S1, S2, …, S6. Subject 3 has been excluded due to low sampling rate which did not allow the calculation of the necessary mean sigma index feature. The Dreams database with two visual scorings is limited to six subjects on which a few existing algorithms have tested their performance.75
Table 4-6 Mass database results obtained for montage 1 (collection of C3, CZ, C4 electrodes)...........77
Table 4-7 Results from all Montages of the Mass Database. SD columns in the table represent standard deviation.............................................................................................................................................................................78
Table 4-8 Mass Database Slow and Fast Spindle Results, two different cut-offs were used. 13 Hz cut-off (De Gennaro and Ferrara, 2003) and 14 Hz cut-off (Dang-Vu et al., 2003). A shift of majority is seen when the cut-off is changed from 13 Hz to 14 Hz. This implies a majority of the spindles scored were between the 13Hz to 14 Hz range. When comparing the frequency distributions of the spindles scored by the algorithm and the spindles scored by visual scorers, no significant differences were observed as the one-way ANOVA showed $p > 0.05$. .................................................................................................................................................................................80
Table 4-9 Results showing differences in False Positive Proportion and $F_1$-score when NREM 3 spindles are not considered in the MASS database (D1) and Charité database (D2). Data from all subjects was combined to obtain the results in this table.................................................................................................................................................................................82
Table 4-10 Results comparing the union of scoring and the intersection of scoring for the MASS database using the C3 channel. Data from all subjects was combined to obtain the results in this table. .................................................................................................................................................................................83
Table 4-11 Results showing the Pearson’s correlation coefficient between the number of spindles scored by the visual scorers versus the number of spindles scored by the automatic detection algorithm for different subjects.................................................................................................................................................................................83
Table 4-12 Results showing spindle density (spindles/minute) for different sleep stages. The mean (standard deviation) have been shown using our method, Visual scoring and Ray et al., [115] using the MASS database.................................................................................................................................................................................84
Table 4-13, Charité database results using the IIR filters method. An overall sensitivity of 65.18% at an FPP of 119.74% was observed.................................................................................................................................................................................85
Table 4-14 Results obtained using the Random Forrest Classifier (MASS database).........................87
Table 4-15 Sensitivity, FPP and F₁-score results for different window sizes ........................................... 88
Table 4-16 Results obtained using Pattern matched wavelet #1. Overall Sensitivity of 84% at a Precision of 62% was observed .................................................................................................................. 89
Table 4-17 Results obtained using pattern matched wavelet #2, Overall Sensitivity of 74% at a Precision of 42.9% was observed .................................................................................................................. 90
Table 4-18, Summary Table showing best results assessed using F₁-score obtained using the different sleep spindle detection methods .................................................................................................. 91
Table 4-19, Summary Table showing best results assessed using F₁-score obtained using the different K-complex detection methods .................................................................................................. 91
Table 6-1 Memory consolidation and SS improvement analysis – Subject 1- auditory stimulation only protocol .......................................................................................................................... 102
Table 6-2 Memory consolidation and SS improvement analysis – Subject 2 - auditory stimulation only protocol .......................................................................................................................... 103
Table 6-3 Memory consolidation and SS improvement analysis – Subject 3 - auditory stimulation + NF protocol ......................................................................................................................... 103
Table 6-4 Memory consolidation and SS improvement analysis – Subject 4 - auditory stimulation + NF protocol (Subject could only do 7 sessions) ......................................................................................... 103
Table 6-5 Memory consolidation and SS improvement analysis – Subject 5 - NF only protocol (Subjects could only do 4 sessions) .................................................................................................... 104
Table 6-6 Memory consolidation and SS improvement analysis – Subject 6 - NF only protocol .................................................................................................................................................. 104
Table 6-7 Post nap questionnaire results. Results show no significant differences from baseline from any of the subjects, these are discussed further in Chapter 7 ........................................................................ 105
Table 6-8 Sleep spindle count and density (spindles per minute) of first three sessions for all subjects. ................................................................................................................................................. 106
Abstract

Sleep spindles and K-complexes are transient events that are observed in electroencephalography (EEG) signals of the mammalian brain. In human sleep, spindles and K-complexes have been positively correlated with significant functions such as memory consolidation and intelligence. This dissertation addresses two issues related to sleep spindles and K-complexes, i.e., one is the use of automatic detection algorithms to detect sleep spindles and K-complexes, and the other is the exploration of auditory stimulation to enhance sleep spindle count during sleep.

The numbers of sleep spindles and K-complexes during sleep vary significantly between different subjects. This implies that automatic detection algorithms need to be developed with sensitivity to such differences. In this dissertation, multiple detection methods are presented. The primary methods presented here use the expectation maximisation (EM) clustering technique in order to be sensitive to inter-subject differences. It was found that clustering certain features using the EM algorithm produced results that were able to capture the aforementioned inter-subject differences. The results for sleep spindle detection were evaluated on two public databases and a private database, with a total of 27 subjects. While simultaneous addressing the problem of inter-subject differences, comparison with existing methods in literature showed that the performance of the EM based clustering technique was on par with what had been reported in the existing literature. The results for K-complex detection were evaluated on one public database with 6 subjects. Compared with existing methods in literature, the K-complex detection algorithm presented in this work showed poorer performance but the algorithm was effective in capturing inter-subject differences. Due to the lack of availability of multiple public K-complex databases, a conclusive finding could not be reached for the use of clustering algorithms to detect the K-complex.

This dissertation also presents results observed and analysed using two different methods to enhance sleep spindles during 90-minute naps. One method used auditory stimulation synchronised to a post spindle refractory period to enhance sleep spindle count. Sleep spindle detection algorithms developed using clustering techniques in the first part of the dissertation were used to synchronise auditory stimulation to a post spindle refractory period. The other method to enhance sleep spindles was to use sensorimotor rhythm (SMR) neurofeedback (NF) prior to a 90-minute nap. Three groups consisting of 2 subjects each were used in this study. Group 1 received auditory stimulation only for 5 sessions, and Group 2 received auditory stimulation and NF for 10 sessions, while Group 3 received NF only for 10 sessions. The results showed that both Group 1 and Group 2 showed an increase in sleep spindles compared with the baseline, while Groups 3 showed negative performance compared with the baseline. Results showed that only group 1 showed an improvement in memory consolidation compared with the baseline, while Groups 2 and 3 showed negative performance compared with the baseline. These results indicate that auditory stimulation synchronised to a post spindle refractory period is effective in enhancing sleep spindles and memory consolidation, while SMR NF showed an inverse relationship in enhancing sleep spindles and memory consolidation.
Overall, the research presented in this dissertation made novel contributions with the application of clustering techniques to sleep transient detection and auditory stimulation synchronised to a post spindle refractory period.
## Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AASM</td>
<td>American Academy of Sleep Medicine</td>
</tr>
<tr>
<td>ANN</td>
<td>Artificial Neural Network</td>
</tr>
<tr>
<td>ANOVA</td>
<td>Analysis Of Variance</td>
</tr>
<tr>
<td>AUC</td>
<td>Area Under Curve</td>
</tr>
<tr>
<td>CWT</td>
<td>Continuous Wavelet Transform</td>
</tr>
<tr>
<td>DWT</td>
<td>Discrete Wavelet Transform</td>
</tr>
<tr>
<td>EEG</td>
<td>Electroencephalography</td>
</tr>
<tr>
<td>EM</td>
<td>Expectation Maximisation</td>
</tr>
<tr>
<td>EMG</td>
<td>Electromyography</td>
</tr>
<tr>
<td>EOG</td>
<td>Electrooculography</td>
</tr>
<tr>
<td>FMRI</td>
<td>Functional Magnetic Resonance Imaging</td>
</tr>
<tr>
<td>FPP</td>
<td>False Positive Proportion</td>
</tr>
<tr>
<td>FT</td>
<td>Fourier Transform</td>
</tr>
<tr>
<td>Hz</td>
<td>Hertz</td>
</tr>
<tr>
<td>IIR</td>
<td>Infinite Impulse Response</td>
</tr>
<tr>
<td>MASS</td>
<td>Montreal Archive of Sleep Studies</td>
</tr>
<tr>
<td>MCI</td>
<td>Memory Consolidation Index</td>
</tr>
<tr>
<td>MGB</td>
<td>Medial Geniculate Body</td>
</tr>
<tr>
<td>MGM</td>
<td>Multivariate Gaussian Model</td>
</tr>
<tr>
<td>MGMMM</td>
<td>Multivariate Gaussian Mixture Model</td>
</tr>
<tr>
<td>NF</td>
<td>Neurofeedback</td>
</tr>
<tr>
<td>NREM</td>
<td>Non Rapid Eye Movement</td>
</tr>
<tr>
<td>PPV</td>
<td>Positive Predictive Value</td>
</tr>
<tr>
<td>PSG</td>
<td>Polysomnography</td>
</tr>
<tr>
<td>Acronym</td>
<td>Definition</td>
</tr>
<tr>
<td>---------</td>
<td>------------------------------------</td>
</tr>
<tr>
<td>REM</td>
<td>Rapid Eye Movement</td>
</tr>
<tr>
<td>RMS</td>
<td>Root Mean Squared</td>
</tr>
<tr>
<td>ROC</td>
<td>Receiver Operator Characteristic</td>
</tr>
<tr>
<td>SMR</td>
<td>Sensorimotor Rhythm</td>
</tr>
<tr>
<td>SPWRS</td>
<td>Sharp-wave Ripples</td>
</tr>
<tr>
<td>SS</td>
<td>Sleep spindles</td>
</tr>
<tr>
<td>STFT</td>
<td>Short-Time Fourier Transform</td>
</tr>
<tr>
<td>SVM</td>
<td>Support Vector Machine</td>
</tr>
</tbody>
</table>
Symbols

$x(t)$  Continuous time signal in $t$

$X(F)$  Fourier transform of $x(t)$

$X_{STFT}(F, t)$  Short time Fourier transform of $x(t)$

$w(t)$  Windowing function centred at $t$

$P_{\sigma}(t)$  Sigma Power feature

$I_{\sigma}(t)$  Sigma Index feature

$P_{\sigma_2}(t)$  Sigma Power 2 feature

$I_{\sigma_2}(t)$  Sigma Index 2 feature

$y[n]$  Discrete time signal with index $n$

$x[n]$  Discrete time signal with index $n$

$a_k$  Reverse coefficients of infinite impulse response filter

$b_k$  Forward coefficients of infinite impulse response filter

$H(z)$  Transfer function of infinite impulse response filter

$\Omega$  Angular frequency

$R_{\sigma}(t)$  Sigma Ratio feature

$M_{\sigma}(t)$  Mean Sigma Index feature

$x$  Vector of features

$\mu$  Vector of feature mean’s

$\Sigma$  Covariance matrix

$\Theta$  Set of all parameters required to define a multivariate Gaussian mixture model

$\alpha_i$  Mixture weight of $i^{th}$ component of Gaussian mixture model

$\theta_i$  Parameters of $i^{th}$ component of Gaussian mixture model

$L()$  Likelihood function

$log(L())$  Log likelihood function

$\chi$  Set of observed data
\(x_i\) \(i^{th}\) vector element of \(X\)

\(Y\) Set of hidden or unobserved data

\(y_{i}\) \(i^{th}\) scalar element of \(Y\)

\(\delta\) Knocker delta function

\(\mathcal{C}\) Set of available classes

\(\mathcal{S}\) Binary set of separations of a decision tree node

\(G\) Gini Index

\(h_k(x)\) Single decision tree

\(\mathcal{H}\) Set of ensemble decision trees

\(G_K\) Slope of K-complex

\(I_{\Delta L}\) Low delta index feature

\(X_{cwt}(t, s)\) Continuous wavelet transform value of a signal at time \(t\) and scaling factor \(s\)

\(\mathcal{F}\) Family of polynomial functions

\(\rho_i\) Polynomial function of degree \(i\)

\(\lambda\) Lagrange multiplier

\(G\) Matrix

\(g_{ij}\) Element from \(i^{th}\) row and \(j^{th}\) column of matrix \(G\)

\(m\) Vector

\(b\) Vector

\(\alpha\) Vector

\(z\) Vector

\(c\) Random variable

\(\mathcal{B}\) Beta function

\(y_{ik}\) \(i^{th}\) sample from \(k^{th}\) group used in ANOVA

\(n_k\) Number of samples in group \(k\)
<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu_i$</td>
<td>Mean of group $i$.</td>
</tr>
<tr>
<td>$\text{Diag}()$</td>
<td>Matrix with diagonal elements</td>
</tr>
<tr>
<td>$B$</td>
<td>Matrix</td>
</tr>
<tr>
<td>$A$</td>
<td>Matrix</td>
</tr>
<tr>
<td>$\text{adj}()$</td>
<td>Adjoint operator/ Adjoint of Matrix</td>
</tr>
<tr>
<td>$\text{tr}()$</td>
<td>Trace operator</td>
</tr>
</tbody>
</table>
List of Publications by the Author


The best bridge between despair and hope is a good night’s sleep.

E. Joseph Cossman
1 Introduction

This dissertation is built on concepts from the three different fields, including signal processing, neuroscience, and psychology, as illustrated by Fig.1-1. The first two chapters explain the fundamental concepts, theory, principles and techniques from these three fields which underpin the research work presented in the successive chapters.

![Diagram of three fields: Signal Processing, Neuroscience - Neural mechanisms of sleep, Psychology - Memory consolidation tests and concepts.]

Figure 1-1 Three fields of studies which underpin the work presented in this dissertation.

1.1 Brief history of sleep research

Sleep has been a primordial necessity for adequate functioning of humans and mammals, and a necessity whose purpose has been researched and questioned since antiquity. Earliest documented theories of sleep have been attributed to the Greek physician Alcmaeon (500 BC) who argued that sleep is a loss of consciousness occurring when blood drains from the vessels on the surface of the body [1][2]. Although many such pseudo-scientific theories of sleep have been put forward through time, a substantial understanding of sleep has only developed in the 20th century with the arrival of modern technologies and modern research methods. In 1929, the German scientist and physician Hans Berger was the first to develop electroencephalography (EEG), a method to record and study the electrical activity from the brain [3]. Using EEG, Berger noticed a difference in brain activity during sleep and wakefulness, and hence paved the way for a modern quantitative research method to understand sleep. Along with EEG, the electroocculogram (EOG) signal captures electrical activity from eye movement and the electromyography (EMG) which produces a signal to represent electrical activity from muscle movements are predominantly studied in modern day sleep research [4].
1.2 Electroencephalography (EEG) Rhythms

Since the first use of EEG by Hans Berger in 1929, a number of different characteristic oscillations in
different frequency ranges were observed by researchers. These rhythms are usually known by their
different Greek letters and are briefly described as follows.

*Alpha band* (8 – 13 Hz) brain waves, also called the Berger waves, were one of the first rhythms to be
discovered. Rhythms in alpha band are hypothesised to represent the idle brain rhythm based on the
fact that subjects with closed eyes show an increase in alpha band activity in the parietal and occipital
regions (corresponding to visual cortex) of the brain. The generation of alpha waves is attributed to
thalamocortical feedback loops [5].

*Beta waves* (14 – 20 Hz) are frequency patterns which are maximally observed over the frontal and
central regions of the scalp and are associated with active thinking or an active cortex [6].

*delta activity* (0.5 – 4 Hz) is predominantly seen during the deep sleep in humans and is characterised
by a very high peak to peak amplitude [4].

*theta activity* (4 – 7 Hz) is hypothesised to facilitate in the formation of new semantic and episodic
memories [7].

Frequency activities above 25 Hz are usually called *gamma activity* in literature [6].

1.3 Sleep Stages and Sleep Transients

1.3.1 Sleep Stages

Sleep has been characterised into different stages based on characteristic patterns observed in EEG,
EOG and EMG. A normal night sleep usually involves a series of cycles going through non-rapid eye
movement (NREM) sleep and rapid eye movement (REM) sleep and back to NREM sleep. These
cycles usually last 90-120 minutes. NREM sleep which takes up the major time share of sleep is
further divided into NREM-1, NREM-2 and NREM-3 stages [4].

Standards have been developed to standardise the identification of different sleep stages.
Rechtschaffen and Kales produced one of the earliest research standard to classify sleep into different
stages [8]. More recent standards for classification of sleep stages were produced by the American
academy of sleep sciences (AASM) in 2007 and 2010 [4][9]. A brief overview of the individual stages
as described by the standards is provided as follows.
1.3.1.1 NREM-1

NREM-1 is the first stage of sleep when a subject starts to fall asleep. This stage lasts for a very short period and is characterised by vertex sharp low amplitude (4-7Hz) EEG waves as shown in Fig. 1-2.

Figure 1-2 Transient EEG events seen in wake, NREM-1 and NREM-2. Adapted from [10].

1.3.1.2 NREM-2

Stage 2 of NREM sleep is characterised by two important transient events observed in EEG – the sleep spindle (SS) (shown in Fig. 1-2) and the K-complex. Not only are the SS and the K-complex two of the most researched patterns in the sleep research, they are also the primary objects of the research in this dissertation. The SS is a short burst of electrical activity in the 12 – 15 Hz frequency range, lasting 0.5 – 2 seconds [4]. The number of SS during a whole night’s sleep can vary between subjects and is usually into the hundreds and thousands. The underlying neural mechanisms that generate spindles are discussed further in Chapter 2. K-complexes are also short lasting EEG patterns (1 – 3 seconds) with high amplitude vertex sharp waves containing both a single negative voltage wave followed by a positive voltage wave [4].

1.3.1.3 NREM-3

The NREM-3 sleep stage is defined by continuous high amplitude slow wave activity observed in EEG (0.5 – 2Hz) [4]. SS are also observed in this stage, however the number and density of SS in NREM-3 is far lower than in NREM-2 stage. It is also characterised by minimal activity in EOG and EMG [4]. Because of the predominant slow waves observed in NREM-3, this stage is also called slow
wave sleep (SWS). NREM-3 is the dominant sleep stage in a whole night recording since it can account for nearly 50% of sleep time.

1.3.1.4 REM sleep

REM sleep as its name implies is characterised by rapid movement of the eyes [4]. This is observed in EOG as rapidly changing high amplitude waves. Very low EMG activity and 2 – 6 Hz EEG activity are also observed in the REM stage. The REM is also popularly known as the dream stage of sleep. Fig. 1-3 shows typical EOG, EEG and EMG signals seen during REM sleep.

![Figure 1-3 Typical pattern of EOG, EEG and EMG signals observed during REM sleep.]

1.3.2 Sleep Transients

1.3.2.1 Clinical importance of sleep transients

This dissertation is primarily concerned with sleep spindles and K-complexes. Sleep spindles and K-complexes are clinically significant since a reduced number of these transients have been observed in a number of psychiatric and sleep disorders such as Alzheimer’s disease [11] and Schizophrenia [12]. Sleep spindles are also highly correlated with memory consolidation and motor sequence learning [13]. A more thorough literature review of current theories of sleep transients and their significance is presented in Chapter 2.

1.3.2.2 Automatic sleep transient detection

Visual identification of sleep transients is a process in which a sleep scoring expert visually identifies sleep transients while looking through sleep EEG data. This is a very time consuming process due to the sheer number of sleep spindles and K-complexes in sleep data. Furthermore, there is significant disagreement between two different sleep experts in scoring the same sleep transients [14]. To counter
this problem, researchers have been testing methods of automatic detection using signal processing and pattern recognition methods [15]. A number of these methods developed are based on features extracted from the EEG signal that are compared with some threshold values to identify the presence of a sleep transient [15][16][17]. Most previous automated sleep transient detection methods have used some form of thresholds to identify sleep transients. The problem with this thresholding based approach is that the characteristics and power of sleep transients differ in subjects. Part of the novel work in this dissertation has been to circumvent the problem associated with thresholds by using clustering methods which have not been applied to sleep transient detection before.

A detailed review of existing methods is presented in Chapter 2.

1.3.2.3 Methods to enhance sleep spindle density

Due to the positive correlation between SS count (number of sleep spindles in a sleep session), memory consolidation and general healthy mental states, researchers have designed methods over the years to enhance SS in EEG of subjects collected in sleep experiment sessions [18][19]. Two methods of enhancing SS count and density are the use of low amplitude auditory stimulation during sleep and neurofeedback prior to sleep. A brief introduction to both methods is given, as follows.

1.3.2.3.1 Auditory stimulation to enhance sleep spindles density

Current research models of sleep indicate that a neural network between the thalamus and the neurons of the cortex are responsible for the generation of SS [20]. SS have been hypothesised by some researchers to function as an inhibitory system that restricts external stimuli from disrupting sleep [22]. Based on this idea, researchers have used auditory stimulation to trigger spindles in the NREM-2 and NREM-3 sleep stages. Research by Sato and colleagues in 2008 has shown that an increased density of spindles is observed in healthy individuals following auditory stimulation (manually triggered) in the NREM-2 sleep stage [19]. Ngo and colleagues used automated auditory stimulation in NREM-3 sleep stage to enhance slow wave sleep and sleep spindles [18]. The second part of this dissertation describes an experimental protocol and the corresponding results obtained using auditory stimulation during sleep. The algorithms needed and developed for that auditory stimulation protocol are described in Chapter 3 of the dissertation. The novelty in this research is the use of automatic auditory stimulation triggered following automatic detection of sleep spindles. This is unlike earlier studies which have triggered automatic auditory stimulation triggered with the detection of slow wave sleep [19].
1.3.2.3.2 Neurofeedback to enhance sleep spindles density

Brain oscillations observed via EEG show a range of frequencies which can change significantly with time, behaviour and the mental faculties currently in use. For example, tasks requiring attention [22] and memory recall [23] can show different power in different frequencies compared with the idle resting state.

Neurofeedback (NF) is a type of biofeedback where participants learn to self-regulate brain activity using real-time feedback of EEG features [24][25]. Using this process enables participants to have real-time feedback of their EEG activity (traditionally shown on a computer screen in the form of power in specific frequency bands), enabling them to voluntarily activate different brain states. One NF protocol which has been experimented with in relation to sleep is the sensorimotor rhythm (SMR) feedback. SMR NF involves enhancement of the 12 – 14 Hz rhythm observed in the sensorimotor cortex (central regions of the scalp). Research by Kober et al., in 2014 [24], and Hodmoser et al., in 2008 [25], has shown SMR NF prior to sleep as being effective in increasing SS count and density during sleep. The second part of the dissertation will show how an SMR NF protocol prior to sleep is implemented to compare with the auditory stimulation protocol which is used as a benchmark during sleep.

1.4 Research Aims

The aims of the research work presented here are twofold. First, devise computer based automatic sleep transient detection algorithms which are sensitive to inter-subject differences. Second, apply these algorithms (specifically the sleep spindle detection algorithms) to test a real-time feedback system which has been designed to enhance SS count and density during sleep.

1.4.1 Research Objectives

The research presented in this dissertation can be broadly categorised into the following research objectives as follows.

1. Develop and test automatic sleep transient detection methods using clustering techniques, including:
   a. Sleep spindle detection using clustering techniques;
2. Implement a real-time auditory stimulation system to enhance sleep SS using the automatic sleep transient detection methods developed in Objective 1.
3. Conduct a pilot study to test the real-time auditory stimulation system to enhance SS along with the SMR NF protocol.

4. Compare the results obtained using the real-time auditory stimulation protocol and the SMR NF protocol.

1.4.2 Hypotheses

Four hypotheses are tested in this research, including that:

1) Automatic detection of SS using clustering techniques without the use of thresholds is effective detection methods to observe inter subject differences.

2) Automatic detection of K-complexes using clustering techniques without the use of thresholds is an effective detection method to observe inter subject differences.

3) SS count can be increased and SS can be entrained with real-time auditory stimulation using automatic SS detection algorithm developed in Objective 1.

4) Increase in SS count using auditory stimulation is higher than increase in sleep spindle count using the SMR NF protocol.

1.4.3 Significance and Limitations:

This section describes the limitations and significance of individual hypothesis in sub-section 1.4.2

1) Testing of hypothesis 1 is done using statistical tests with significance levels of 0.05, i.e., a 5% confidence interval. Using an algorithm that uses the 5% confidence interval would imply that the algorithm is practical to use instead of visual scorers to compare inter-subject differences. The testing of this hypothesis is limited by a medium sized sample (25 subjects across two databases).

2) Testing of hypothesis 2 follows similar practical significance, technical significance and limitations compared to hypothesis 1. Using an algorithm that uses the 5% confidence interval would imply that the algorithm is practical to use instead of visual scorers to compare inter-subject differences. The testing of this hypothesis was limited to a small sized sample of 6 subjects due to the lack of a larger online database for K-complex detection.

3) Testing of hypothesis 3 was undertaken using a pilot study with a limited number (6) subjects across multiple sleep sessions. The ANOVA statistical test with a technical significance level of 0.05% was used to test this hypothesis. The practical significance at the 5% confidence interval would mean that SS can be entrained with real-time auditory stimulation using automatic SS detection algorithm developed in Objective 1.
4) Testing of hypothesis 4 was undertaken using the same pilot study as used for hypothesis 3 with a limited number (6) subjects across multiple sleep sessions. The ANOVA statistical test with a technical significance level of 0.05% was used to test this hypothesis.

1.5 Contributions of Thesis Work

During the course of the research conducted as presented in this dissertation, a number of contributions to research have been made. These contributions along with the hypotheses they answer (previous section) are listed here:

1) Novel methods of automatic sleep spindle detection using clustering techniques have been developed, introduced and evaluated. These methods were aimed at answering hypotheses 1 and 2. Novel features such as Sigma Power 2 and Sigma index 2 (sub-section 3.2.2.1) using IIR filters were also developed during this process.

2) A novel method of auditory stimulation has been developed and tested in order to increase SS count during day-time naps. This method is aimed at studying hypotheses 2 and 4.

3) A comparison of two protocols, i.e., auditory stimulation during sleep and SMR NF before sleep has been conducted to check their efficacy in increasing SS and improving sleep induced memory consolidation. This comparison helped in testing hypothesis 4.

4) Two other novel methods (not involving clustering techniques) of sleep transient detection were developed and tested. These are:
   a. An SS detection method using the Random Forest Classifier; and
1.6 Thesis Outline-

Chapter 1 presents a basic introduction to the research and lays the foundation to understand the following chapters. Research aims and working hypotheses are also presented in this chapter.

Chapter 2 expands on Chapter 1 by discussing current literature in the field of sleep transient research and NF research. This chapter borrows from the three different research areas of signal processing and pattern recognition, neuroscience of sleep spindle generation, and psychology (clinical significance of sleep transients). A detailed literature review of existing automatic sleep transient detection research is presented along with literature detailing state of the art neuroscience and mathematical models of SS generation. This is followed by a discussion of current literature on the clinical significance of sleep spindles.

Chapter 3 describes the signal processing methods developed and the databases used to test the automatic transient detection methods.

Chapter 4 presents the results of methods described in Chapter 3.

Chapter 5 describes the methods of the pilot study involving auditory stimulation and NF to enhance SS count.

Chapter 6 presents the results of the pilot study described in Chapter 5.

Chapter 7 discusses the results presented in Chapters 4 and 6 with respect to existing literature and provides conclusions including closing statements with regard to the hypotheses stated in Chapter 1.

Relevant Appendices are also presented. These appendices help understand core mathematical concepts described in earlier chapters and are cited in relevant sections where necessary.
2 Sleep Spindle and K-complex definitions, detection methods, generation mechanisms and significance

This chapter provides a review existing definitions, automatic sleep transient detection methods, neural mechanisms of sleep transient generation and existing understanding of their functional significance in humans. The concepts and literature discussed in this chapter set a platform to understand the following chapters which discuss the research conducted during the course of this dissertation.

2.1 Sleep Spindle and K-complex Definitions

2.1.1 American Academy of Sleep Medicine definitions

The AASM provided the latest definition of a sleep spindle and K-complex in 2010. Sleep spindle is defined as a “train of distinct waves with frequency 11-16Hz (most commonly 12-14 Hz) with a duration ≥ 0.5 seconds, usually maximal in amplitude using central derivations” [4] (Fig. 1-1). K-complex is defined as a “well-delineated negative sharp wave immediately followed by a positive component standing out from the background EEG, with total duration ≥ 0.5 seconds, usually maximal in amplitude when recorded using frontal derivations” [4], as shown in Fig. 2-1.

![Figure 2-1 Typical K-complex showing a well delineated negative sharp wave immediately followed by a positive component. Maximal amplitude of a K-complex is usually observed using the frontal derivations, i.e. the signals from electrodes placed towards the front of the scalp.](image)
2.1.2 Definitions by Rechtschaffen & Kales

The Rechtschaffen and Kales [8] sleep classification published in 1968 has a much narrower definition of SS. They are defined as 12-4 Hz rhythmic bursts for at least 0.5 seconds.

The K-complex is defined as a “well delineated negative sharp wave immediately followed by a positive component”. A minimum duration of 0.5 seconds is also set. This definition of a K-complex is similar to the definition from the AASM sleep classification standard.

2.2 Automatic sleep spindle and K-complex detection methods

2.2.1 Automatic sleep spindle detection

A number of fully automated SS detection algorithms have been developed over the years. They can broadly be divided into methods based on machine learning such as neural networks and methods based on time-frequency analysis of the EEG signals. The following sub-sections describe some significant methods developed by researchers in the field and the databases used to test such methods.

2.2.1.1 Online databases

Prior to 2011, all SS detection algorithms developed by researchers were tested on their own private sleep databases since there had been no common/public database available. As a consequence, the results obtained by different researchers were not directly comparable. In 2011, Devuyst et al., [26] published a public database called the DREAMS SS database consisting of six 30-minute sleep excerpts. The SS in those excerpts were visually scored by two different sleep experts. The publication of this database meant that researchers could test their algorithms on this database and consequently compare with other methods tested on the same database. A much larger database called the MASS (Montreal Archive of Sleep Studies) database was published by the Centre for Advanced Research in Sleep Medicine in 2014 [27]. This database consisted of overnight sleep data collected from 19 subjects. The SS in that data were also scored by two different sleep experts. The larger MASS database with longer overnight recordings provided a comprehensive database to test SS detection methods. Most new SS research since 2014 had tested their SS detection methods on the MASS database [16][17].
2.2.1.2 Accuracy measurement parameters

To measure the performance of an algorithm, researchers have used traditional measures such as sensitivity, specificity, accuracy, precision and F1-score. Details of these measures are provided in Section 3.1 (Performance Measures).

2.2.1.3 Bandpass filtering- thresholding methods for SS detection

One of the earliest methods of SS detection was developed by Schimicek and colleagues [28], where the signal was band passed using an infinite impulse response (IIR) filter in the SS frequency range followed by detection based on a set threshold. Fig. 2-2 shows a basic illustration of this method.

![Illustration of a thresholding detection method.](image_url)

Figure 2-2 Illustration of a thresholding detection method. An EEG signal (top plot) is filtered in the SS frequency band; this is followed by identifying filtered signal segments whose power is greater than a set threshold (bottom plot).

Physical movement of a subject during an EEG recording produces EMG signals due to muscle movements. Since EMG signals exhibit power in a broad spectrum, they effect EEG recordings across multiple frequencies. These interference's are known as EMG artifacts and they are known to affect EEG recordings in the SS frequency band [28].
In Schimicek’s method, EMG artifacts in EEG were removed if the power in a 5 second EEG signal was more than 5uV². Since Schimicek’s algorithm, other automatic detection methods were developed based on similar principles of band pass filtering. In 2000, Huupponen et al., [29] developed a spindle detection algorithm based on fuzzy detection and amplitude threshold technique. In the first stage of their process, Huupponen et al., detect sleep spindles using a fuzzy detector with a pre-specified threshold. In the second stage of the process, an amplitude threshold in the 10-16Hz range is recalculated using spindles detected in the first stage. This was done using a Bayesian approach to maximise the number of spindle detected. The algorithm was tested on 4 subjects (using a private database) with a sensitivity ranging from 73.4% to 83.1% and a false positive rate from 2.3% to 5.4%. It can be said that the processes used by Huupponen et al., [29] was semi-adaptable to individual subjects. In 2007, Huupponen et al., developed another method which introduced a new feature called the sigma index [14]. This feature is the ratio of power in the 10.5-16 Hz range to the sum of powers in the 4-10 Hz (Theta + Low Alpha frequency) and the 20-40 Hz bands. The sigma index based detector also consisted of two stages. In the first stage, a fixed amplitude threshold of 4.5 was used to detect SS. In the second stage, the mean amplitude of the previously detected SS in the first stage was used to set a new threshold. The method was tested on a private database of 12 subjects, and produced a sensitivity of 70.6% at a specificity of 98.6%.

In 2010 Ray et al., also developed an automated detection method based on band pass filtering and threshold detection. However, the threshold used was based on first 12 visually identified spindles (by an expert) from an individual subject. A sensitivity of 98.96% at a specificity of 88.49% was obtained.

2.2.1.4 Artificial Neural Networks

Artificial neural networks (ANN) have been used for SS detection. ANN is a supervised learning classifier which requires the use of training data to train the classifier. Spectral features obtained from the short-time Fourier transform (STFT) of 0.5 second EEG signals were used to feed an ANN in [31], where testing was done on individually selected 1142 equally distributed spindle and non-spindle samples obtained from a private database. They reported an accuracy of 88.7% on the testing set. Features extracted from the Autoregressive Model of EEG samples were used as input features to an ANN in [32], where the ANN was trained on a private database containing data from 12 subjects, and data from 6 subjects was used for testing. A Sensitivity of 89.1% at a False detection rate of 9.3% was reported. An EEG signal samples filtered in the 10.5-16 Hz range were used as the inputs to an
ANN in [33], which was tested on a single subject and showed a sensitivity of 92.9% at a specificity of 86.9%.

### 2.2.1.5 Wavelet transform methods

Use of wavelet transform based methods has been quite popular with researchers in the past. Features extracted by using a Daubechies 4 wavelet filter along with features extracted by using the Teager energy operator were used to detect SS in [34], where the algorithm was tested on 95 individually obtained spindles and showed an overall accuracy of 93.9%.

The discrete wavelet transform (DWT) was used as part of a larger decision tree based SS detection algorithm in [35], where features extracted using the DWT were used in some nodes of their decision tree algorithm. The algorithm was tested on data from 16 different sleep recordings of a private database and showed a 96.17% sensitivity at a 95.54% specificity.

More recently, the continuous wavelet transform (CWT) was used to develop a SS detection method in [17], which was tested on the publically available MASS and Dreams databases. Results from the MASS database showed a sensitivity of 84% at a specificity of 90% and a false discovery rate of 83%. Results from the Dreams database showed a sensitivity of 76% at a specificity of 92% and a false discovery rate of 67%.

The thresholding technique with features obtained from the CWT of EEG signals was used to detect SS [36], which was tested on a private database consisting of data from 9 subjects and showed a Sensitivity of 85.4% at an FDR of 86.2%.

The continuous wavelet transform with an individualised threshold based on mean background activity in the 6-18 Hz range was used to develop an SS detection method in [37] which was tested on a private database consisting of data from 49 pairs of twins. A sensitivity of 72%, specificity of 90% at a precision of 40% was obtained.

### 2.2.1.6 Matching pursuit methods

Matching pursuit decompose a signal into a best set of predefined dictionary atoms (or predefined signal patterns). Durka and Blinowski were the first to use the matching pursuit technique to develop an SS detection algorithm [38]. However they did not test their algorithm on an overnight sleep recording or a sleep excerpt. The algorithm was minimally tested on four individual spindles to analyse frequency and amplitude characteristics. Another matching pursuit method was also
developed for SS detection [39], which was tested on 9 overnight sleep recordings and showed a sensitivity and specificity of 81.2%.

2.2.1.7 Other methods

The Teager energy operator and the spectral edge frequency feature with fixed thresholds were used to identify spindles in [40], where the algorithm was tested on the Dreams database and produced a sensitivity of 80.3% with a specificity of 97.6%.

A bivariate normal model of frequency and amplitude was reported to detect SS in [41], which was tested on a private database (with 7 subjects) and the Dreams database. A sensitivity of 71.1% at a specificity of 98.6% was observed on the Dreams database and a sensitivity of 78.5% at a specificity of 94.2% was observed for the private database.

The tunable Q-factor wavelet transform and morphological component analysis were used to detect SS in [42], where the algorithm was tested on the MASS database. Results obtained showed a sensitivity of 83.18% and false discovery rate (FDR) of 39%.

The sparse low-rank optimisation technique with multiple EEG channels was used to detect SS in [43], which was tested on Dreams and The MASS databases. Results obtained on the Dreams database showed a sensitivity of 63%, F_1-score of 66% and a precision of 69%. Results obtained on the MASS database showed a sensitivity of 61%, F_1-score of 62% and a precision of 64%.

Table 2-1 provides a performance overview of major automated sleep spindle detection techniques in terms of sensitivity, specificity, accuracy, precision, F_1-score, false discovery rate, false detection rate, and false positive rate.
Table 2-1 Overview of major research works in automated sleep spindle detection.

<table>
<thead>
<tr>
<th>Publication</th>
<th>Methods Used</th>
<th>Number of Subjects</th>
<th>Database</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Other Measures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Schimicek et al., 1994</td>
<td>Band Pass Filtering followed by thresholding</td>
<td>1</td>
<td>Private</td>
<td>90%</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td>Huupponean et al., 2000</td>
<td>Fuzzy detection and amplitude threshold</td>
<td>6</td>
<td>Private</td>
<td>73.4%-83.1%</td>
<td>NA</td>
<td>False Positive rate – 2.3% to 5.4%</td>
</tr>
<tr>
<td>Huupponean et al., 2007</td>
<td>First stage – Amplitude threshold, Second stage – adaptive threshold based on first stage</td>
<td>12</td>
<td>Private</td>
<td>70%</td>
<td>98.6%</td>
<td>False positive rate – 32%</td>
</tr>
<tr>
<td>Ray et al., 2010</td>
<td>Band pass filtering and thresholding. Threshold set by first twelve visually detected spindles</td>
<td>10</td>
<td>Private</td>
<td>98.96%</td>
<td>88.49%</td>
<td></td>
</tr>
<tr>
<td>Gorur et al., 2002</td>
<td>Features extracted using STFT were classified using a ANN</td>
<td>&lt;1 (tested on individual segments of spindles and non spindles)</td>
<td>Private</td>
<td>NA</td>
<td>NA</td>
<td>Accuracy – 88.7%</td>
</tr>
<tr>
<td>Acir and Guzelis 2004</td>
<td>ANN</td>
<td>6</td>
<td>Private</td>
<td>89.1%</td>
<td>NA</td>
<td>False detection rate – 9.3%</td>
</tr>
<tr>
<td>Ventouras et al., 2005</td>
<td>Band pass filtered EEG signal (10.5-16 Hz) fed into an ANN</td>
<td>1</td>
<td>Private</td>
<td>92.9%</td>
<td>86.9%</td>
<td></td>
</tr>
<tr>
<td>Ahmed et al., 2009</td>
<td>Thresholds on features extracted using</td>
<td>&lt;1 (tested on individual)</td>
<td>Private</td>
<td>NA</td>
<td>NA</td>
<td>Accuracy – 93.9%</td>
</tr>
<tr>
<td>Study</td>
<td>Methodology</td>
<td>Dataset</td>
<td>Sensitivity</td>
<td>Specificity</td>
<td>Positive Predictive Value</td>
<td>Discovery Rate</td>
</tr>
<tr>
<td>------------------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>--------------------</td>
<td>-------------</td>
<td>-------------</td>
<td>---------------------------</td>
<td>----------------</td>
</tr>
<tr>
<td>Duman et al., 2009</td>
<td>Discrete Wavelet Transform, Decision Tree</td>
<td>16</td>
<td>96.17%</td>
<td>95.54%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tsanas and Clifford</td>
<td>Continuous Wavelet Transform.</td>
<td>20</td>
<td>84% for MASS database, Dreams Database</td>
<td>90% for MASS, 92% for Dreams</td>
<td>False discovery rate – 83% for MASS and 67% for Dreams</td>
<td></td>
</tr>
<tr>
<td>O’Reilly et al., 2015</td>
<td>Thresholding features obtained using CWT</td>
<td>9</td>
<td>85.4%</td>
<td>NA</td>
<td></td>
<td>False Discovery rate – 86.2%</td>
</tr>
<tr>
<td>Adamczyk et al., 2015</td>
<td>CWT with individualized threshold.</td>
<td>98</td>
<td>72%</td>
<td>90%</td>
<td></td>
<td>Precision – 40%</td>
</tr>
<tr>
<td>Schonwald et al., 2006</td>
<td>Matching Pursuit</td>
<td>9</td>
<td>81.2%</td>
<td>81.2%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Imtiaz et al., 2013</td>
<td>Teager energy operator and spectral edge frequency features with thresholds</td>
<td>6</td>
<td>80.3%</td>
<td>97.6%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nonclerq et al., 2013</td>
<td>Bivariate Normal model of frequency and amplitude.</td>
<td>13</td>
<td>78.5% for private database and Dreams Database</td>
<td>94.2% for private database and 98.6% for Dreams</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lajnef et al., 2015</td>
<td>Tunable Q-factor Wavelet transform</td>
<td>14</td>
<td>83.18%</td>
<td>NA</td>
<td>False discovery rate – 39%</td>
<td></td>
</tr>
<tr>
<td>Parekh et al., 2017</td>
<td>Sparse low-rank optimization</td>
<td>20</td>
<td>61% for MASS database, Dreams Database</td>
<td>63% for Dreams</td>
<td>F1-score of 62% for MASS and 66% for Dreams</td>
<td></td>
</tr>
</tbody>
</table>
2.2.2 Automatic K-complex detection

Automated K-complex detection is a similar problem to automated SS detection. Both K-complexes and SS are hall marks of the NREM-2 sleep stage and they both present a similar problem in terms of visual detection by experts. However, past research on automated K-complex detection has not been explored as widely as that on automated SS detection. Here, important research findings are reviewed in automated K-complex detection research.

2.2.2.1 Online databases for K-complex detection

Similar to SS databases, prior to 2011 there was no public database for K-complex detection. The Dreams K-complex database consisting of ten – 30 minute excerpts from overnight recordings of different subjects was published. The K-complexes in the excerpts were scored by two different human scores for five of the ten subjects. K-complexes for the other five subjects were only scored by one subject. As such researchers in the past have only evaluated their algorithms on the five subjects which have visual scorings from two experts.

2.2.2.2 K-complex detection methods

One of the first automatic detection methods (including hardware system) was developed based on bandpass filtering and thresholding the peak to peak amplitude [44]. A peak to peak amplitude of 100 μV in a 1 second moving window was used to detect K-complexes in a bandpass filtered (1.1 to 2.3 Hz) signal. Additional constraints such as peak to peak polarity were also set. The system was tested on four different sleep EEG recordings and the K-complexes were visually annotated by three different scorers of which a majority had to agree on a K-complex. Being one of the first studies, only raw results in terms of agreed number of K-complexes between system and scorers were provided. No measure of sensitivity or specificity was provided.

A K-complex detector was developed using ANN and amplitude features (2 second moving window) derived from a single EEG signal in [45], which obtained sensitivity of 90% at a false positive rate of 8% when tested on a test set consisting of 51 K-complexes and 49 non K-complexes.

A continuous density hidden Markov model was used to detect K-complexes in [46], where no result was reported in the traditional performance measures of sensitivity, precision or accuracy. A private
database was used and the model was trained on 1000 K-complexes which were visually identified by 4 different expert scorers.

Fuzzy thresholds with amplitude and power features extracted from filtered EEG signals were used to detect K-complexes in [47]. The training data used consisted of 2 whole night recordings and the testing data was the publically available Dreams database. The results showed a true positive rate of 61.72% at a false positive proportion of 19.62% compared with those by visual scorer 1 and a true positive rate of 60.94% at a false positive proportion of 181.25% when compared with those by visual scorer 2. This study was the first to analyse results on the publically available Dreams K-complex database.

Non smooth optimisation along with a range of classifiers such as support vector machine (SVM) and random forest was used to detect K-complexes in [48], where K-complexes extracted from a private database was used to train and test the classifiers. A training set with 28 non-K-complexes and 31 K-complexes (i.e., a total of 59 observations), and a test set with 38 non-K-complexes and 35 K-complexes (i.e., a total of 73 observations) were used. Best result reported in the form of accuracy was 66.1%.

Features extracted using the wavelet transform and the Teager energy operator were used to detect K-complexes in [49]. Fixed thresholds were used in automatic detection process. Results were obtained on a private database (on 1 subject) with visual scoring of K-complexes performed by three different expert scorers. True positives rates for the three scorers were 80%, 89% and 87%, respectively, corresponding to false positive rates of 7%, 7% and 8%.

Different machine learning methods were used to detect K-complexes in [50]. However, this study was limited to testing 326 segments of EEG, each of 10 seconds duration, among which 54 of the segments were K-complexes and 272 segments were non K-complexes. Best results showed a sensitivity of 70.4% at a precision of 70.4% and an accuracy of 90.2%.

The DWT was used to extract various features to identify K-complexes in [51]. Feature thresholds were used to separate K-complexes from non K-complexes. The algorithm was tested on the publically available Dreams database and showed a mean sensitivity of 74% at a positive predictive value of 65%. Results were also analysed on a second private database with similar sensitivity and positive predictive value.
Convex optimisation based methods was used for K-complex detection in [52], where very small training data set (with 30 non-K-complexes and 21 K-complexes) and a test data set (with 9 non-K-complexes and 10 K-complexes) were used. Convex optimisation performed involved decomposition of an EEG signal into polynomial functions, upon which a classifier was used to train and test for accuracy of detection. The highest accuracy of 84% was obtained using a SVM classifier trained using the sequential minimisation algorithm. The dataset used in this study was a very small private dataset.

A K-complex detection method using Q-factor wavelet transform was presented in [53]. The method used the tunable Q-factor wavelet transform which allows adaptation of a wavelet to the desired transient signal, in this case the K-complex. A training set was used which consisted of 630 sleep EEG data segments in total, including 420 S2 segments and 210 non-S2 segments. Results produced (on 14 subjects overnight data) showed a sensitivity of 80.23% at an false detection rate (FDR) of 37.27% compared with expert scorer 1 and a sensitivity of 82.5% at an FDR of 38% when compared with expert scorer 2.

Similar to the research problems encountered in past SS detection methods, K-complex detection methods in existing literature traditionally use fixed thresholds and failed to address the problem of inter-subject differences. This thesis work has set to address these problems in research with the use of clustering techniques.

Table 2-2 provides a performance overview of major automated K-complex detection techniques in terms of sensitivity, specificity, and ‘other measures’ such as accuracy, precision, false detection rate, and false positive rate.
<table>
<thead>
<tr>
<th>Publication</th>
<th>Methods Used</th>
<th>Number of Subjects</th>
<th>Database</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Other Measures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bankman et al., 1992</td>
<td>Amplitude features with ANN</td>
<td>&lt;1 (51 individual K-complexes and 49 non K-complexes)</td>
<td>Private</td>
<td>90%</td>
<td>NA</td>
<td>False Positive Rate (FPR) – 8%</td>
</tr>
<tr>
<td>Devuyst et al., 2010</td>
<td>Fuzzy thresholds with amplitude and power features</td>
<td>5</td>
<td>Dreams Database</td>
<td>61.72% for scorer 1 and 60.94% for scorer 2</td>
<td>NA</td>
<td>FPR of 19.62 % for scorer 1 and 181.25% for scorer 2</td>
</tr>
<tr>
<td>Moloney et al., 2011</td>
<td>Non smooth optimisation with SVM classifier and the random forrest classifier.</td>
<td>&lt;1 (35 individual K-complexes and 38 non K-complexes)</td>
<td>Private</td>
<td>NA</td>
<td>NA</td>
<td>66.1%</td>
</tr>
<tr>
<td>Erdamer et al., 2012</td>
<td>Wavelet Transform + Teager Energy operator with fixed thresholds</td>
<td>1</td>
<td>Private</td>
<td>89%</td>
<td>NA</td>
<td>FPR of 7%</td>
</tr>
<tr>
<td>Vu et al., 2012</td>
<td>Multiple machine learning methods</td>
<td>&lt;1 (326 segments of EEG each 10 seconds long)</td>
<td>Private</td>
<td>70.4%</td>
<td>NA</td>
<td>Precision – 70.4%, Accuracy – 90.2%</td>
</tr>
<tr>
<td>Krohne et al., 2014</td>
<td>DWT</td>
<td>5</td>
<td>Dreams Database</td>
<td>74%</td>
<td>NA</td>
<td>Precision – 65%</td>
</tr>
<tr>
<td>Zamir et al., 2015</td>
<td>Conver optimization with SVM</td>
<td>&lt;1 (10 individual K-complexes and 9 non K-complexes)</td>
<td>Private</td>
<td>NA</td>
<td>NA</td>
<td>Accuracy – 84%</td>
</tr>
<tr>
<td>Lajnef et al., 2017</td>
<td>Q-factor Wavelets</td>
<td>&lt;1 (630 sleep EEG segments)</td>
<td>Private</td>
<td>80.3% for scorer 1 and 82.5% for scorer 2</td>
<td>NA</td>
<td>FDR of 37.37% for scorer 1 and 38% for scorer 2</td>
</tr>
</tbody>
</table>
2.3 Sleep spindle generation mechanisms

2.3.1 Neural mechanisms
The underlying neural pathways and mechanisms that produce sleep spindles have been studied and a few hypothesis and theories have been suggested. In 1945 Morison and Bassett [54] showed that SS are produced within the thalamus (Fig. 2-3). They observed that SS were still recorded in the thalamus after decortication and removal (transection) of the brain stem (medulla oblongata, pons and mid brain, see image below). Further, in 1974 Villablanca showed that SS are observed even after the removal of the striatum (part of the brain connected to the thalamus and the cerebral cortex) and the rhinencephalon (another part of the brain). In 1985 Jones et al., showed that the Reticular Nucleus (RE) is a thin sheet of GABAergic neurons that cover parts of the thalamus [56]. They receive dendritic inputs from the cerebral cortex and the dorsal thalamic nuclei. The outputs of the RE are connected to the thalamus. Pare et al., showed that SS are absent when observed in parts of the thalamus that have no afferent connections to the RE thereby indicating that the RE is a vital component for SS generation [57].

Figure 2-3 One hemisphere of the brain viewed laterally. Adapted from Sharma and Majsak [125].

Steraiide et al., showed that the SS are generated via closed loop activity between the following three regions: the thalamic cortex – RE – the cerebral cortex (Pyramid cells) [58]. Fig. 2-4 shows an illustration of the spindle generation circuit. Based on in vivo and in vitro experiments [58], cortical pyramid cells receive excitatory inputs from the thalamic nuclei and in turn output excitatory potentials to the reticular thalamic nucleus. The reticular neurons project inhibitory outputs to the
thalamus and also receive excitatory inputs from the thalamus. Low threshold Ca\(^{2+}\) spikes in reticular cells are known to generate 7 to 14 Hz bursts of activity. These bursts in turn produce inhibitory postsynaptic potentials (IPSPs) in the thalamocortical neurons. Since there are also excitatory outputs connecting the thalamocortical neurons to the reticular cells, the IPSP’s inactivate the Ca\(^{2+}\) spikes which are followed by a rebound Ca\(^{2+}\) spike with 7 to 14 Hz bursts. These periodic bursts are transferred to the cortex via IPSP’s from the Thalamocortical neurons. The resulting periodic bursts in the cortex are observed as the SS in EEG.

![Image](image.png)

Figure 2-4 Sleep spindle generating neural circuit showing the Reticular nucleus - Thalamocortical cells - Cortical Pyramid cells loop. A ‘+’ sign indicates excitatory afferents and a ‘−’ sign indicates inhibitory afferents. Modified from Steraide et al. [59].

### 2.3.2 Mathematical Model of generation and intra spindle period

Zygierewicz et al., developed a mathematical model of SS generation using lumped models of the thalamocortical neurons and reticular nucleus [60]. Lumped models as shown in Fig. 2-5 use individual element models for different types of neurons. In their model, lumped elements for inhibitory transmission of GABA neurons and excitatory transmission of the AMPA neurons were used. Since the thalamus is considered a routing hub for the transmission of sensory information [61][62][63], external noise was also parameterised into their model (shown as N(t) in Fig. 2-5). The results of their simulations showed that the average intra-spindle period (time between two successive spindles) was reduced in the presence of external noise. The intra-spindle period based on a TC – RE parameter gain of 3 was six seconds. Zygierewicz et al., showed that the TC – RE coupling parameter is inversely proportional to intra-spindle period.
Koupparis et al., observed that the recurrence rate of sporadic SS suggests that the reduction of power at 1 - 3 seconds (since last spindle occurrence) reflected a minimum refractory period for spindles lasting about 2 seconds [64]. Other researchers have reached similar conclusions in the past regarding the refractory period of spindles. Kim et al., observed that the 3 - 20 second period between individual spindle waves was due to a refractory period during which the threshold for external stimuli to produce another spindle is higher [65]. Intracellular recordings from thalamocortical neurons of a Ferret by Bal and Mccromick in 1996 showed an intra-spindle period between 5 – 20 seconds indicating a minimum refractory period between spindles. In-vitro investigations by Luthi et al [66]. It is shown in [67] that the increases in Ca\(^{2+}\) due to rebound Ca\(^{2+}\) bursts in the reticular nucleus are the cause of the refractory period between SS.

As seen in later chapters, the refractory period between consecutive spindles, also called the intra-spindle period, was the target of the auditory stimulation during sleep in this research investigation. In this research, auditory stimulation during the refractory period/intra-spindle period was hypothesised to increase the total number of sleep spindles during sleep (as described under hypotheses in Chapter 1).
2.4 Functional significance of sleep spindles

Yamadori et al., were amongst the first researchers to study the effects of applying auditory stimulation during sleep [68]. Based on their observation that auditory stimulation applied out of phase with spindles elicited a K-complex whereas auditory stimulation in phase with spindles produced no evoked responses, it was suggested that spindles correlate with neural mechanisms which protect from environmental noise to secure ongoing sleep deepening process.

Experiments on cats in 1981 by Livingstone and Hubel showed that degree of information transfer from the lateral geniculate nucleus (retinal information) to the cortex is based on the state of arousal [69]. Information transfer degraded during periods of drowsiness and sleep showing a possibly role of sleep and sleep spindles in vetoing information transfer.

Using the advanced FMRI (functional magnetic resonance imaging) technology of the 21st century, Dang Vu et al., performed imaging studies of the auditory cortex during periods of auditory stimulation in sleeping subjects [70]. They showed that auditory stimulation out of phase with SS produces a response in the auditory cortex, while stimulation in phase with the SS showed no such activation, further substantiating a sleep protective role of SS during sleep.

In a review article in 2014, Luthi et al., discuss multiple findings from literature including the previously mentioned work of Vu et al. [62][70]. Based on existing literature, Luthi and colleagues conclude that “Spindles separate sleep into distinct phases of sensory accessibility, enabling periods of almost complete annihilation of sensory throughput with periods in which sensory information does reach the cortex. At the same time, spindles enable epochs of heightened plasticity that fall together with the emergence of hippocampal rhythms”. They also went on to support another theory on spindles enhancing memory consolidation by supporting the transfer of information from short-term memory (hippocampus) to long-term memory (cortex).

2.4.1 Sleep spindles, hippocampus ripples and memory consolidation

Sharp-wave ripples (SPWRs) are high frequency (100 – 300 Hz) bursts produced by the hippocampus (area of the brain responsible for short term memory) during quiet wakefulness and NREM sleep [71]. Temporal correlations have previously been observed between SPWRs and SS in rats and humans [72]. This correlation along with experiments on rats and humans has led many researchers to suggest
that SS and slow oscillations during sleep along with SPWRs mediate the transfer of hippocampus memories to long term cortical storage sites [72][73].

2.5 Clinical significance of sleep spindles in memory consolidation
Study of SS following declarative learning tasks prior to sleep showed an increase in SS number and density (in subsequent sleep) based on the study by Gais et al., in 2002 [74]. The learning protocol needed the subjects to memorise lists of word pairs which were later tested by cued recall where the first word of each pair was presented and the second word had to be recalled. Their results showed that when compared with the subjects who did not undertake any declarative learning tasks prior to sleep, the subjects who did showed an increase in sleep-density during the following sleep session.

Schabus et al., examined SS activity in the NREM2 sleep stage following a word-pair association memory task prior to sleep and after waking up [75]. They observed that increased NREM-2 SS activity is related to increase in memory recall performance and hence SS density may reflect memory consolidation.

Based on the above findings, a number of researchers have attempted to alter SS density during sleep, which will be discussed further in section 2.7.1.

2.6 Sleep spindles and K-complexes in abnormal conditions
SS and K-complex count, i.e., the number of total SS and K-complexes during overnight sleep has been shown to be correlated with certain abnormal sleep and psychiatric disorders. In 2007, Ferrarelli et al., observed that the density and number of SS in schizophrenia were significantly reduced and thereby suggested that deficits in the thalamic reticular nucleus and thalamo-reticular circuits in schizophrenia as indicated by sleep spindle count may represent a biological marker for the illness [76].

Wamsley et al., found similar results in terms of SS density and number when comparing schizophrenia subjects with normal subjects. Further they tested the relationship between memory consolidation, SS and schizophrenia, suggesting that abnormal spindle generation impairs sleep-dependent memory consolidation in schizophrenia [77]. Wamsley and colleagues also go on to
suggest that treating sleep spindle abnormalities is a novel target for the treatment of cognitive deficits in schizophrenia.

Reduced SS have also been observed in other abnormal conditions such as Alzheimer’s disease and Supranuclear Palsy [11], while increased density and number of sleep spindles has also been hypothesized to be an indication of superior cognitive function and intelligence [78] [11].

A significant part of sleep transient research in the past has focused on SS while a few studies have focused on K-complexes and their characteristics in diseased subjects. In one such study, Ramakrishnan et al., studied K-complex number and density in schizophrenia while measuring problem solving performance [79]. The findings suggest that density of K-complexes is directly proportional to problem solving ability in schizophrenia. Many past hypotheses of the function of SS have pointed to a sleep protective role along with their part in memory consolidation, and a similar sleep protective role has been suggested for K-complexes. Forget et al., studied auditory stimulation during sleep in controls and insomnia subjects and observed that it affected both spontaneous K-complexes and evoked K-complexes (evoked by external stimuli) [80]. Based on increased density of evoked K-complexes following auditory stimuli, their findings supported a sleep protective role of K-complexes.

Given the correlations of sleep transient’s density with healthy cognitive function, attempts have been made in designing and testing methods to enhance SS and K-complex count and density. These are presented in the following sections.

2.7 Sleep spindle and K-complex enhancement methods

2.7.1 Auditory stimulation to enhance sleep spindles and K-complexes

Early research developments in the study of auditory pathways of the brain were studied in animals such as cats and Guinea pigs. In 1983, Calford and Aitkin were among the first researchers to study auditory pathways from the cochlear nuclei to the cortex in cats [81]. Their findings suggested that there existed multiple parallel auditory pathways that connect cochlear nuclei to the cortex; going through the thalamus. Studying auditory pathways in Guinea pigs in 1995, King et al., observed auditory evoked responses produced by two different stimuli presented at random in series [82]. They observed that the EEG responses to changing stimuli were prominent at EEG locations that showed no response when the same stimuli was presented in series. Based on their findings, they suggested that primary and non-primary auditory pathways provide distinct contributions to encoding auditory
stimuli. Multiple parallel pathways connecting the cochlear nuclei to the auditory cortex have been researched and documented by other researchers as well such as research conducted on cats by Imig and Morel [83].

The relay of acoustic information to the human/mammalian cortex is performed by two distinct pathways: the lemniscal and non-lemniscal neurons [84]. The lemniscal pathways are located in the ventral parts of the medial geniculate body (MGB) which is a part of the thalamus responsible for relaying auditory information. Unlike the lemniscal pathways, the non-lemniscal pathways are situated in the dorsal and medial parts of the MGB. The lemniscal pathways are known to be sensitive to stimulus distinct from the background whereas the non-lemniscal pathways are described to supply information about background noise or environmental changes [85][86][87]. Researchers have shown that the non-lemniscal pathways are desensitised to repetitive stimulation using the same auditory frequencies [88].

Based on a review of existing literature, Bellesi and colleagues made recommendations that suggest the use of randomly varying acoustic stimulus for enhancing sleep spindles and K-complexes [87]. This is to prevent habituation on the non-lemniscal pathways to the auditory cortex whereby ensuring a response to auditory stimuli during sleep.

In 2007 Sato et al., studied the effects of sensory stimulation in the NREM2 sleep stage [19]. Effects of somatosensory, auditory or visual stimulation were studied in different subjects. In order to study the effects of auditory stimulation, Sato and colleagues presented repeated auditory stimulation (1Khz tone, varying duration 1-3 seconds) for 5 minutes in NREM-2 and compared the stimulation period with a non-stimulation period. Their results showed a significant increase in SS number during the stimulation periods and the observed latency from onset of stimulus to succeeding spindle was an average of 2 seconds. The minimum latency between successive spindles was supportive of earlier theories of a minimum refractory period between successive spindles [66][67].

Ngo et al., showed that auditory stimulation in phase with the slow oscillations in NREM3 increased spindle and slow oscillation density [18]. In a stimulation group, auditory stimulation with changing frequency (pink noise) and duration of 50 milliseconds was presented to subjects in phase with the positive peaks of slow oscillations. In the sham group (i.e., control group), no auditory stimuli were presented. The changing frequency auditory stimulation was in line with recommendations made by Bellesi et al. [87]. Comparison of the two groups showed a significant increase in SS and slow
oscillation synchronisation, count and density. This is one of the first studies showing entrainment of sleep transients using auditory stimulation. Their research also studied the effects of memory consolidation due to auditory stimulation. Using a declarative memory test (word-pair association) it was shown that the auditory stimulation in phase with slow oscillations improved memory consolidation significantly.

In 2016, Ong et al., studied auditory stimulation in NREM-2 and NREM-3 sleep stages [89]. Auditory tones were presented throughout a nap session of 90 minutes in the stimulation condition and no auditory stimulation in the sham condition. Declarative memory consolidation was also studied between the two groups. Pink noise bursts of 50 millisecond duration were used for auditory stimulation. In agreement with the work done by Ngo et al., (2014) [18], improved memory consolidation was observed in the stimulation group along with increase in fast spindle activity and slow wave activity.

Antony et al., presented repeated bursts of white noise oscillation to subjects undertaking 90 minute naps [90]. Participants were assigned to three different conditions, a 12 Hz stimulation condition, a 15 Hz stimulation condition and a 50 Hz stimulation condition. In each condition, the participants were presented with white noise and frequency modulated white noise (i.e., the 12 Hz, 15 Hz or 50Hz modulation). Participants were presented with repeating 2 seconds on, 8 seconds off iterations of the group specific corresponding modulated sound starting in the NREM-2 sleep stage. It was shown that the 12 Hz and 15 Hz modulated stimulus conditions showed an increase in fast spindle activity and slow oscillations.

2.7.2 Neurofeedback to enhance sleep spindles

As introduced in chapter 1, NF is a form of biofeedback in which participants are allowed to learn to manipulate their brain activity using feedback parameters provided by EEG or other such modalities (e.g., FMRI, MEG). EEG NF is the modality described in this dissertation as an EEG NF protocol was used in one experiment. SMR NF as introduced in chapter 1 involves the use of the EEG power in the 12-15 Hz band as the feedback parameter. Participants undertaking this form of feedback are required to learn to increase the SMR frequency power in EEG thereby modulating activity in the sensorimotor cortex. SMR NF has in the past been shown to increase sleep spindle density and corresponding overnight memory consolidation.
In 2008, Hoedlmoser et al., conducted an experiment in which they studied the effects of SMR NF prior to sleep on subsequent sleep parameters [25]. 27 healthy subjects (13 male) were divided into two groups – a SMR conditioning group and a control group. Subjects were required to undertake a declarative memory task before and after a 90-minute nap session. While the subjects in the experimental group were asked to enhance amplitude in the SMR frequency range, the subjects in the control group were provided a random frequency range between 7 and 20 Hz. A needle on a computer screen depicting the band power in the corresponding frequency was shown to participants. The needle was designed and programmed to move to the left with an increase in the power in the frequency band of interest, i.e., the 12-15 Hz band for the experimental group. Rewards in the form of audio-visual cues (“appearance of a sun for 2 seconds accompanied by a 200 ms lasting sound of 800 Hz”) were presented to participants if they successfully moved the needle beyond a pre-set threshold based on baseline recordings. Successful conditioning in the SMR band was reported after 10 sessions by each participant in the experimental group. A reported increase in sleep spindles and declarative memory consolidation was also observed along with shortened sleep onset latency.

In 2014, Schabus et al., studied the effects of SMR feedback prior to sleep in subjects with insomnia [91]. 24 subjects with primary insomnia participated in the study. The study was counterbalanced so that each subject participated in SMR training sessions and sham-conditioned training sessions. Similar to the previously described study by Hoedlmoser et al., Schabus and colleagues reported an increase in the 12-15 Hz EEG activity over the course of 10 SMR sessions. They also showed an increase in subjective sleep quality, slow wave sleep, overnight memory consolidation, sleep spindles, and a decrease in number of awakenings in the SMR group.

Kober et al., in 2015 also conducted a similar SMR NF study to that by Hoedlmoser et al., [25]. An experimental group of 10 subjects was used along with a control group consisting of an equal number of subjects. While the experimental group received SMR feedback, the sham group received random inputs as feedback. Unlike the needle on a computer screen described by Hoedlmoser et al., Kober et al., used bars the move up and down to show the respective increase or decrease in SMR activity. Along with a central bar showing the 12-15 Hz SMR activity, two side bars representing 4 - 7 Hz activity and 20-30 Hz activity were also presented. Subjects were required to voluntarily learn to move the central bar up, while keeping the side bars down. The side bars were representative of eye movement artifacts (4-7 Hz) and muscle movement artifacts (20-30 Hz). Both artifacts are known to show a false increase in the 12-15 Hz SMR band. The findings of this study included and observed
linear increase in SMR power over training runs, which was associated with improvements in memory, attention and a corresponding increase in SS density.

In 2017, Schabus et al., conducted a double-blind study [92] based on a similar single blind study performed by them in 2014 [91]. The study examined the effects of SMR NF prior to sleep on sleep parameters in subjects with insomnia. A placebo feedback group and a SMR NF group were used in the experiment. Unlike other studies this study did not show any improvements in memory consolidation or sleep spindle parameters.
3 Automated sleep transient detection – Fundamentals

This chapter describes the methods that were developed for automatic sleep transient detection. Although the major developments in this research have been the application of clustering techniques to sleep transient detection, other methods have also been developed to allow a comparison of results. All methods developed during the candidature are described in the following sections. Pseudo codes for these methods are described in Appendix D.

3.1 Performance Measures

The ‘By-Event’ measurement standard described by Warby et al., has been used to calculate ‘True Positives’, ‘False Positives’ and ‘False Negatives’ [15]. Visual scoring from two individual scorers was available for Dreams and MASS databases. The union of the two scorers for each database was used to calculate the measures. Using the ‘By-Event’ standard, any overlap between an SS detected automatically and a visually scored SS were considered as True Positives.

The parameters estimated include [15][26]:

\[ T_p – \text{True Positives}, \ F_n – \text{False Negatives}, \ F_p – \text{False Positives}, \text{and} \ T_n – \text{True Negatives}. \]

\[ \text{Sensitivity (Recall)} = \frac{T_p}{T_p+F_n} \]  

(1)

The sensitivity (or recall) measure describes the ratio of the number of true spindles that were detected by both the algorithm and visual scoring \((T_p)\) to the total number of spindles scored by visual scoring \((T_p+F_n)\).

\[ \text{False Positive Proportion} = \frac{F_p}{T_p+F_n} \]  

(2)

The false positive proportion is a measure that describes the ratio of the number of false positives i.e., the number of spindles detected by the algorithm but not by visual scoring, to the total number of spindles scored by visual scoring.

\[ \text{Specificity} = \frac{T_n}{T_n+F_p} \]  

(3)

Specificity measure is a ratio of the number of true ‘non-spindles’ i.e., segments or samples of signal that are not part of a spindle which are agreed upon by both the algorithm and visual scoring \((T_n)\) to the total number of non-spindles scored by the visual scoring \((T_n+F_p)\).

\[ \text{Precision} = \frac{T_p}{T_p+F_p} \]  

(4)

Precision measure describes the ratio of the number of true spindles that were detected by both the algorithm and visual scoring \((T_p)\) to the total number of spindles scored by the algorithm \((T_p+F_p)\).
Note the difference between sensitivity and precision. The sensitivity measure compares $T_p$ to the total number of spindles scored by visual scoring, while the precision measure compares $T_p$ to the total number of spindles scored by the algorithm.

$$F_1 = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$$

(5)

An ideal algorithm would need high sensitivity (recall) along with high precision. Since there is a usual trade-off between sensitivity and precision for most non-ideal algorithms, the $F_1$-score measure describes the harmonic mean between sensitivity and precision. This allows for a comparison between multiple algorithms through this one statistic, i.e., the $F_1$-score statistic.

3.2 Automated sleep spindle detection

3.2.1 Databases

Results were analysed on three different databases that contained visual SS scorings. Two of these were public databases and one was a private database. The details of the databases are discussed as follows.

3.2.1.1 Dreams Database – 1

The polysomnography (PSG – combination of EEG, EOG, EMG and respiration signals) sleep data (of Dreams sleep spindle database) containing spindles were obtained from the University of MONS - TCTS Laboratory (Stéphanie Devuyst, Thierry Dutoit) and Université Libre de Bruxelles - CHU de Charleroi Sleep Laboratory (Myriam Kerkhofs) online database [26]. The database contains six 30-minute sleep excerpts extracted from whole night recordings of 3 male and 3 female subjects aged between 30 and 55 years. Visual scoring of spindles was undertaken by two independent scorers and the scoring data were available as part of the Dreams database. This thesis work introduced automated scoring methods which were compared with the visual scoring results carried out by the two experts on the online database. The PSG data consisted of three EEG, two EOG and one EMG channels. PSG data for subject numbers 2, 4, 5 and 6 were recorded at a sampling rate of 200Hz. PSG data for subject 1 was recorded at 100 Hz and data for subject 3 was at 50Hz.

3.2.1.2 MASS Database – 2

The SS2 cohort (one of the PSG data sets of many PSG sets) database from the Montreal Archive of Sleep Studies (MASS) was used in [27]. The database consists of data from 19 subjects, including 8 male and 11 female subjects. The age of participants ranged between 18 and 33 years. All EEG channels were sampled at 256 Hz. Visual scoring was performed by two different scorers named ‘V4’ and ‘V5’. Whist V4 used AASM rules for scoring, V5 used a broader definition and a wider band (11-17Hz) to score spindles [30]. Union of scores by V4 and V5 were used to compare results in this study. Data from subjects 4, 8, 15 and 16 were excluded from analysis since sleep spindle scores by scorer V5 were not available for these subjects.

3.2.1.3 Charité Database – 3

This private database consisted of overnight sleep PSG data obtained from six (including four male and two female) healthy human subjects. Only EEG signals sampled at 200 Hz were utilized in this
analysis. Sleep spindles were visually scored by a single expert. This data was recorded in a sleep lab at the Universitätsmedizin Berlin: Charité (Berlin, Germany). Unlike the Dreams database and the MASS database, this is a private database.

3.2.2 Clustering technique
The clustering technique developed during this candidature is based on features extracted using the STFT or infinite impulse response (IIR) filters. Once the features were extracted, they were then clustered as Gaussian mixture models (GMM) [93] using the expectation maximisation (EM) algorithm [94]. A pre-processing stage similar to the method used in [28] was used to remove EMG artifacts from data. The following sections describe the features extracted using two different methods followed by a description of the EM algorithm.

3.2.2.1 Features Using Short-Time Fourier Transform
In the first study of this research investigation, features were extracted using the STFT and classified using the EM algorithm.

The Fourier transform (FT) of a signal is useful in extracting frequency specific information in signals and is defined as [95][97][98]:

$$X(F) = \int_{-\infty}^{\infty} x(t) e^{-i2\pi F t} \, dt$$  \hspace{1cm} (6)

where \( x(t) \) and \( X(F) \) represent the signal as a function of time, \( t \), and its Fourier transform as a function of frequency, \( F \), respectively.

The FT provides overall frequency information from a given signal. To obtain more temporally specific frequency information, the signal can be split into many segments and the Fourier transform calculated for individual segments. This leads to what is commonly known as the short-time Fourier transform (STFT) as defined in (7) [95][96][97][98].

$$X_{STFT}(F, t) = \int_{-\infty}^{\infty} x(\tau) w(\tau - t) e^{-i2\pi F \tau} \, d\tau$$  \hspace{1cm} (7)

where \( w(\tau - t) \) is the windowing function which restricts the Fourier transform to a specific time segment.

Four spindle related features were extracted and then classified using the EM approach. To extract the four features, an STFT moving window of a given length was used. During analysis of results, this moving window length was changed over a range of values to examine the effect of varying window length on the results. An (N-1) overlap moving window was used, i.e., the moving window was shifted by each sample. For purposes of demonstration and explanation, a 1.5 second moving window size (with no zero padding) is used in the following descriptions. In the following equations, assume that the features are being calculated for an input signal, \( x(t) \).
For each window, the following features were calculated:

1) **Sigma Power:**

\[
P_\sigma(t) = \int_{12.5}^{15} X_{STFT}(F, t) dF
\]

(8)

where \( P_\sigma(t) \) is defined as a function of time \( t \), and \( X_{STFT}(F, t) \) represents the STFT of \( x(t) \), as defined in (7), as a function of frequency \( f \) and time \( t \). \( P_\sigma(t) \) is a direct representation of the power in the SS frequency range.

2) **Sigma Index:**

\[
I_\sigma(t) = \frac{\int_{12.5}^{15} X_{STFT}(F, t) dF}{\int_{0}^{40} X_{STFT}(F, t) dF}
\]

(9)

where \( I_\sigma(t) \) is defined as a function of time \( t \), and \( X_{STFT}(F, t) \). \( I_\sigma(t) \) is a representation of the ratio of power in the SS frequency to power in the major EEG frequency range of 0.5 – 40 Hz [26]. This feature was previously defined and used by Devuyst et al. [26]. Since artifacts from other sources such as EOG and EMG affect the complete EEG spectrum, this provides a feature that differentiates artifacts from spindles.

3) **Sigma Power 2:**

\[
P_{\sigma_2}(t) = \frac{P_\sigma(t)}{P_\sigma(t-1.5)+P_\sigma(t+1.5)}
\]

(10)

where \( P_{\sigma_2}(t) \) is calculated as the ratio of sigma power \( P_\sigma(t) \) of a 1.5 second window (in this case, if the window length was changed, so was the adjacent window length) to the spindle power in the two adjacent moving windows to the current window. This feature is a novel contribution to research. This feature is illustrated in Fig. 3-1. Since spindles are limited to the 0.5–2 second time length, it is assumed that if a spindle exists in a moving window, there are no spindles in the adjacent moving windows (i.e., spindles are separated by the length of the moving window, as shown in Fig. 3-1). A high \( P_{\sigma_2}(t) \) indicates a high probability of an SS existing in the current moving window. Since SS are
transient (limited to a maximum 2.0 seconds), this feature is developed to penalise activity in the SS frequency range lasting for longer durations.

4) **Sigma Index 2**: 
\[
I_{\sigma_2}(t) = \frac{I_{\sigma}(t)}{I_{\sigma}(t-1.5) + I_{\sigma}(t+1.5)}
\]  

(11)

where \( I_{\sigma_2}(t) \) is defined in a similar way to \( P_{\sigma_2}(t) \). It is the ratio of the relative spindle power in a moving window to the sum of relative spindle powers in the two adjacent moving windows. Similar to \( P_{\sigma_2}(t) \), this feature was developed (novel feature) to support the transient nature of SS.
Figure 3-1 Signal showing the four STFT features along with the original signal (top plot). A spindle is detected in the region marked ‘Spindle region’.

The four features are shown in Fig. 3-1 for a 15 second epoch. The top plot in Fig. 3-1 shows a signal with a 0.5 second spindle represented using the rectangular red box which is also the moving window used to calculate the STFT features. The green boxes are the adjacent windows of the interest. The sigma power and sigma index features show a high amplitude in the red box area, i.e., the spindle
region as there is high SS frequency band activity there. Since there is not SS frequency band activity in the adjacent windows regions, there is low amplitude shown in the sigma power and sigma index. Further since the sigma power 2 and sigma index 2 features are comparing red area of the sigma power and sigma index features to the adjacent green areas, these sigma power 2 and sigma index 2 features also show high amplitude in the spindle region.

3.2.2.2 Features Using Infinite Impulse Response Filters

In the previous section, a description of features developed using the STFT method was described. In this section features developed using the IIR filters are described. Although the features from both methods were eventually classified using the same EM algorithm, the approaches were considered as two separate methods and the results are also presented separately in later chapters.

3.2.2.2.1 IIR Filters

Filters in signal processing are tools in software or hardware that are used to extract or suppress information from a signal. Digital filters are filters that are applicable to discretised data. The two categories of digital filters are the finite impulse response (FIR) and the infinite impulse response (IIR) filters [95][97]. A digital filter applied to an input sequence will provide an output sequence of data which has been altered in frequency and phase characteristics from the input sequence. An IIR filter is defined by the following equation [95][97][98]:

\[ y[n] = \sum_{k=0}^{M} b_k x[n - k] - \sum_{k=1}^{N} a_k y[n - k] \]  

(12)

where \( b_j \) are forward coefficients and \( a_k \) are reverse coefficients. The \( N, M \) summation indexes define the number of \( a_k \) and \( b_j \) coefficients and hence the order of the feedback filter and the feedforward filter, respectively. The general IIR filter direct form transfer function is given by the following equation [97],

\[ H(z) = \frac{\sum_{k=0}^{M} b_k z^{-k}}{1 + \sum_{k=1}^{N} a_k z^{-k}} \]  

(13)

The frequency and phase response of an IIR filter is characterised by its transfer function. Many different forms of IIR filters such as the Chebyshev, Butterworth and Elliptic filters exist in literature [95][97][98]. The Butterworth IIR filter is a filter type known to have a smooth response at all pass band frequencies [97][98] and all filters used in this dissertation were Butterworth filters. The magnitude square function of the Butterworth low pass filter in the analogue domain is given by [97][98],

\[ |H(j\Omega)|^2 = \frac{1}{1 + \left( \frac{\Omega}{\Omega_0} \right)^{2N}} \]  

(14)
where $\Omega_0$ is the -3 db cutoff point of the low pass filter and $N$ is the order of the butterworth filter. For a given $N$ and $\Omega_0$, it is possible to calculate the poles of (14) in the analogue domain and use the bilinear transform approximation to obtain digital coefficients of the Butterworth IIR filter from the respective analogue representations [97].

3.2.2.2.2 Pre-processing to remove EMG artifacts.
Based on the algorithm developed in [28], EMG artifacts were removed using 5-second epochs of the root mean squared (RMS) power in the 20-40Hz band. Epochs with RMS power greater than 5uV^2 were discarded.

3.2.2.2.3 Features
An electrode montage signal is a weighted average signal of a collection of individual EEG channels from different locations on the scalp. Different EEG electrode montages from international 10-20 system [99] were used to compare spindle detection results. Since the availability of the specific channel signals from the three databases were different, only central electrodes were used from the Dreams ($C_z$ only) and Charité databases ($C_3$, $C_4$). For the MASS database, montages were developed using the central, parietal and frontal electrodes. All channels used were monopolar. Since SS are known to be highly synchronous across scalp sites, montages can be used for SS detection [36].

![EEG electrode positions in the 10-20 system, adapted from [99].](image)

The montages are listed as follows:
Dreams database –
Central – $C_z$

MASS database –
$M_1$ – Central – ($C_1$, $C_z$, $C_4$)
$M_2$ – Central + Parietal + Frontal – ($F_7$, $F_8$, $C_z$, $C_3$, $C_4$, $P_3$, $P_4$, $P_5$)
\[ M_c - \text{Central + Frontal} - (F_1, \ F_2, \ C_1, \ C_2, \ C_3) \]
\[ M_p - \text{Central + Parietal} - (C_1, \ C_2, \ C_3, \ P_1, \ P_2, \ P_3) \]

Database 3 –
Central – \( C_1, \ C_2 \)

The resulting signal from each montage was a linearly weighted signal from the combination of channels, e.g., \((C_1 + C_2)/2\) was used for Database 1. The signal obtained from a montage is represented as \( S(t) \) (the weighted average signal from a collection of EEG channels) in the following sections.

The \( S(t) \) signal was filtered using 8th order Butterworth IIR filters to obtain the following signals.

\[
\begin{align*}
B_1(t) &: S(t) \text{ filtered in the 10.5-15Hz band;} \\
B_3(t) &: S(t) \text{ filtered in the 4-10Hz band;} \\
B_4(t) &: S(t) \text{ filtered in the 20-40Hz band;}
\end{align*}
\]

The band frequencies represent spindle band (\( B_1 \)), alpha band (\( B_3 \)) + theta band (\( B_3 \)) and beta band (\( B_4 \)).

**Features**

A moving window of length 1 second was used to calculate the features.

**Sigma Ratio** – This feature is calculated by a two-step process. First, the ratio of RMS power in \( B_1(t) \) to RMS power in \( S(t) \) is calculated. This is ratio is denoted as \( R(t) \) . Second, the ratio of \( R(t) \) to its adjacent windows is calculated as sigma ratio

\[
R_{\sigma}(t) = \frac{R(t)}{(R(t-1)+R(t+1))}
\]

where \( t \) is the time variable in second.

Sigma ratio feature addresses two factors in determining a spindle. It is a ratio of spindle band energy to total energy in the signal. Additionally, it also compares the ratio to similar ratios in adjacent windows. Since high alpha activity can intrude into the spindle band (10.5-15Hz), comparing the spindle band ratio to its neighbouring areas ensures that this feature is only high when a spindle is localised unlike alpha EEG which last for longer durations.

**Mean Sigma Index** – This feature was an adaptation of a feature developed by Huupponen et al. [14]. The feature produces high values when the energy in the spindle band is high and the energy in the alpha EEG band and high beta band is low, since high alpha EEG can intrude into spindle activity and high beta EEG band activity is an indication of EMG artifacts. This feature is defined as follows,

\[
M_{\sigma}(t) = \frac{\mu(|B_1(t)|)}{\mu(|B_2(t)|)+\mu(|B_4(t)|)}
\]
where $\mu(\cdot)$ is a mean function, $|\cdot|$ is the function returning absolute value, $B_i(t)$, for $i = 1, 3, 4$, implies signal epoch in a moving window centred at time $t$.

### 3.2.2.3 Expectation Maximisation Algorithm

The input features were clustered into two sets. Set 1 consists of non-spindle segments and Set 2 of spindle segments. The clustering was based on the expectation maximization (EM) algorithm [94]. In order to separate spindles from non-spindles, multivariate Gaussian mixture model (MGMM) approach was used to cluster the features into a spindle and non-spindle clusters. In this study the EM algorithm was applied to the probability density function (PDF) of a multivariate Gaussian model (MGM) which is defined as follows [100],

$$p(x) = \frac{1}{\sqrt{2\pi^d|\Sigma|}} e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)}$$  \hspace{1cm} (17)

where $d$ is the dimension of feature vector $x = [x_1 \ x_2 \ \ldots \ x_d]^T$ with $x_i \in \mathcal{R}$ for $1 \leq i \leq d, T$ as the matrix/vector transpose operator, $\mu$ is a vector of $d$-dimensions consisting of mean of individual features, and $\Sigma$ is the covariance matrix of the features.

An MGMM is defined as follows [100],

$$p(x|\theta) = \sum_{i=1}^{M} \alpha_i p_i(x|\theta_i)$$  \hspace{1cm} (18)

where

$$\theta = \{\alpha_i, \ldots, \alpha_M, \theta_i, \ldots, \theta_M\},$$  \hspace{1cm} (19)

$\alpha_i$ are the mixture weights for individual $p_i$, $\theta_i$ are the required parameters to define $p_i$, i.e., $\mu_i, \Sigma_i$, and individual $p_i$ are probability density functions defined by (17).

Eq (18) described a probability distribution which is a combination of multiple component Gaussian distributions. Each component Gaussian distribution is defined by its own mean and covariance matrix along with a weighting factor $\alpha_i$.

Further the sum of mixture weight is equal to one, i.e.,

$$\sum_{i=1}^{M} \alpha_i = 1$$  \hspace{1cm} (20)

The likelihood function for the MGMM (18) is given by [100],

$$L(\theta|\mathcal{X}) = \prod_{i=1}^{N} p(x_i|\theta)$$  \hspace{1cm} (21)

where $\mathcal{X} = \{x_1, x_2, \ldots, x_N\}$ are the set of observed data.

The likelihood function is an indicator of how aptly the data fit the model defined by $\theta$. Sometimes it is easier to work with log-likelihood rather than likelihood, which is defined in (22).
From (21), the log-likelihood is defined as follows,
\[
\log(L(\theta|X)) = \log \prod_{i=1}^{N} p(x_i|\theta) = \sum_{i=1}^{N} \log \left( \sum_{j=1}^{M} \alpha_j p_j(x_i|\theta_j) \right)
\] (22)

The above equation is analytically difficult to maximise due to the logarithm of the summation on the right hand side. Expectation maximisation provides a solution for such problems using an iterative method and by introducing hidden or unobserved data/variables in (22).

Assume a hidden data set \( \mathcal{Y} = \{y_1, ..., y_N\} \) corresponding to each \( x_i \). In the case of the MGMM, the \( y_i \) are assumed to provide information as to which mixture component generated the corresponding \( x_i \). Therefore given \( M \) mixture components, \( y_i \in \{1, ..., M\} \).

In this study, there were only two clusters for the spindle detection problem and the K-complex detection problem, i.e., a spindle cluster and a non-spindle cluster for the spindle detection problem and a K-complex cluster and a non-K-complex cluster for the K-complex detection problem. Therefore, \( y_i \in \{1, 2\} \). However, as a generalisation of the equations, \( y_i \in \{1, ..., M\} \) is used in the following equations.

Re-writing the log-likelihood function using the hidden variables along with the original data [100],
\[
\log(L(\theta|X, \mathcal{Y})) = \log(P(X, \mathcal{Y}|\theta)) = \sum_{i=1}^{N} \log(P(x_i|y_i) P(y_i)) = \sum_{i=1}^{N} \log \left( \alpha_{y_i} p_{y_i}(x_i|\theta_{y_i}) \right)
\] (23)

The EM algorithm is divided into two steps. In the first step, the expected value of the log-likelihood function is estimated. The second step involves optimising the parameters to maximise the expected value of the log-likelihood function in the first step.

An initial guess of \( \Theta = \Theta^0 \) is required to initiate the process. In this study, random samples drawn from the observation data were used to calculate initial \( \theta_{y_i}^{0} \) and the mixing proportions \( \alpha_{y_i}^{0} \) were uniform (since there are only two clusters in this case, \( \alpha_{y_i}^{0} = 0.5 \)).

Expected value of complete data likelihood can be calculated by varying the hidden variable –
\[
Q(\Theta, \Theta^0) = \sum_{\mathbf{y}} \log(L(\Theta|X, \mathbf{y})) p(\mathbf{y}|X, \Theta^0)
\] (24)

where \( \mathbf{y} = [y_1, y_2, ..., y_N]^T \) is an instance of the hidden variable, independently drawn from \( \mathcal{Y} \) (the space of \( \mathbf{y} \)).

The probability of \( \mathbf{y} \) in (24) is given by [100] -
\[
p(\mathbf{y}|X, \Theta^0) = \prod_{i=1}^{N} p(y_i|x_i, \Theta^0)
\] (25)

The initial mixture proportions \( \alpha_{y_i}^{0} \) can be thought of the prior probabilities of the mixture components i.e., \( \alpha_{y_i}^{0} = p(y_i|\Theta^0) \), therefore it follows using Bayes rule that [100]

\[
p_{y_i}(y_i|x_i, \Theta^0) = \frac{p(y_i|\Theta^0)p_{y_i}(x_i|\Theta^0)}{p(x_i|\Theta^0)} = \frac{\alpha_{y_i}^{0} p_{y_i}(x_i|\Theta^0)}{p(x_i|\Theta^0)}
\] (26)
Substituting (26) into (24) yields

$$Q(\Theta, \Theta^g) = \sum_{y \in \mathcal{Y}} \left[ \log(L(\Theta | X, y)) \prod_{j=1}^{N} p(y_j | x_j, \Theta^g) \right]$$

(27)

Substituting (23) into (27) yields

$$Q(\Theta, \Theta^g) = \sum_{y \in \mathcal{Y}} \left( \sum_{i=1}^{M} \sum_{y_{i-1}}^{N} \log(\alpha_i p_i(x_i | \theta_{y_{i-1}})) \prod_{j=1}^{N} p(y_j | x_j, \Theta^g) \right)$$

(28)

$$= \sum_{y_{i-1}}^{M} \sum_{y_{i-1}}^{N} \sum_{i=1}^{N} \log(\alpha_i p_i(x_i | \theta_{y_{i-1}})) \prod_{j=1}^{N} p(y_j | x_j, \Theta^g)$$

(29)

$$= \sum_{y_{i-1}}^{M} \sum_{y_{i-1}}^{N} \sum_{i=1}^{N} \sum_{l=1}^{M} \delta_{l,y_{i-1}} \log(\alpha_i p_l(x_i | \theta_{l})) \prod_{j=1}^{N} p(y_j | x_j, \Theta^g)$$

(30)

where $\delta$ is the Knocker delta function and $l$ is a summation index.

Eq. (30) can be simplified to the following equation, (see Appendix for details)

$$Q(\Theta, \Theta^g) = \sum_{i=1}^{M} \sum_{l=1}^{N} \log(\alpha_i p_l(x_i | \theta_{l})) p(l|x_i, \Theta^g)$$

(31)

$$= \sum_{i=1}^{N} \sum_{l=1}^{M} \log(\alpha_i) p(l|x_i, \Theta^g) + \sum_{l=1}^{M} \sum_{i=1}^{N} \log(p_l(x_i | \theta_{l})) p(l|x_i, \Theta^g)$$

(32)

In the maximisation step, the derivatives of (32) are set to zero to find the updated parameters $\Theta$.

Maximizing mixture weights $\alpha_i$, requires the use of a Lagrange multiplier $\lambda$ due to the constraint $\sum_{i=1}^{M} \alpha_i = 1$. This leads to the following [100]:

$$\frac{\partial}{\partial \alpha_i} \left( \sum_{i=1}^{M} \sum_{l=1}^{N} \log(\alpha_i) p(l|x_i, \Theta^g) + \lambda \left( \sum_{i=1}^{M} \alpha_i - 1 \right) \right) = 0$$

(33)

or

$$\sum_{i=1}^{M} \left( \sum_{l=1}^{N} \frac{1}{\alpha_i} p(l|x_i, \Theta^g) + \lambda \right) = 0$$

(34)

$$\sum_{i=1}^{N} \sum_{l=1}^{M} \frac{1}{\alpha_i} p(l|x_i, \Theta^g) + \sum_{i=1}^{M} \lambda = 0$$

(35)

Multiplying both sides of (35) with $\alpha_i$ yields

$$\sum_{i=1}^{N} \sum_{l=1}^{M} p(l|x_i, \Theta^g) + \sum_{i=1}^{M} \lambda \alpha_i = 0$$

(36)

or

$$\sum_{i=1}^{N} 1 + \lambda \sum_{i=1}^{M} \alpha_i = 0$$

(37)

$$\lambda = -N$$

(38)

Substituting $\lambda$ back into (34) yields

$$\alpha_i^{\text{new}} = \frac{1}{N} \sum_{i=1}^{N} p(l|x_i, \Theta^g)$$

(39)
\( \alpha_t^{\text{new}} \) from (39) are the updated mixture weights that can be used in the next iteration.

The update equations for \( \theta_t \), i.e., the mean (\( \mu \)) and covariance matrix (\( \Sigma \)) in the case of MGMM are given here (see appendix A for derivations) as follows,

\[
\mu_t^{\text{new}} = \frac{\sum_{i=1}^{N} x_i p(l | x_i, \theta_t)}{\sum_{i=1}^{N} p(l | x_i, \theta_t)} \quad (40)
\]

\[
\Sigma_t^{\text{new}} = \frac{\sum_{i=1}^{N} p(l | x_i, \theta_t)(x_i - \mu_t^{\text{new}})(x_i - \mu_t^{\text{new}})^T}{\sum_{i=1}^{N} p(l | x_i, \theta_t)} \quad (41)
\]

The new parameters in (39), (40) and (41) are used for the expectation step of the next iteration of the process. In this study, the iterations stopped once the convergence reached a value below 0.000001, i.e., \( L(\theta^{\text{new}} | \mathcal{X}) - L(\theta^{\text{old}} | \mathcal{X}) < 0.000001 \). The EM algorithm has been analytically proved to converge [94][104].

### 3.2.2.4 Spindle Length Check to Separate False Detections

Based on the AASM manual [4], spindles are required to have a minimum length of 0.5 seconds. Since the clustering algorithms which were developed in this study used moving windows, a 0.5 second length spindle could last the entire moving window length. False spindles can be removed by setting a minimum length post detection of spindles by the clustering algorithm. This minimum length was established empirically by varying the length and checking for results. This variation leads to a receiver operator characteristic curve (ROC) as presented in the results section [95][26]. Ideal results can be obtained by using an intersecting line connecting the top left corner and bottom right corner of the ROC curve plots [95][26]. The intersection points give ideal trade-off between Sensitivity and FPP measures.

### 3.2.3 Random Forest Classifier

#### 3.2.3.1 Random Forest Classifier

Random forest classifiers are a type of ensemble classifiers where a number of weaker sub-classifiers (decision trees) are used to develop a stronger classifier. The use of multiple trees and random feature sub-sets is used to reduce the problem of over-fitting to training datasets [105]. A brief overview of decision trees and random forest classifiers is provided here.

#### 3.2.3.1.1 Decision trees

A decision tree is a classification tree in which each node splits the input data into further sub-nodes based on a specific feature threshold. Nodes are added to the decision tree until all data (training set) in the final classification belongs to one class. The following figure shows a simple decision tree which used three features to split input data into two different classes.
3.2.3.1.2 Gini Criterion [106]

The feature at each node on the decision tree along with the threshold (in cases of classification) can be based on a few different information gain measures. In this research, the Gini criterion or Gini index was used to build the decision trees.

Consider the set of input data \( \mathcal{X} = \{x_1, x_2, \ldots, x_n\} \), where \( x_i \in \mathbb{R}^{d \times 1} \) belonging to classes in \( \mathcal{C} = \{c_1, c_2, \ldots, c_k\} \).

The data at every node in a classification decision tree is split into two sets in \( \mathcal{S} = \{\mathcal{S}_1, \mathcal{S}_2\} \) and the probability of individual sets is defined as follows,

\[
P(\mathcal{S}_j) = \frac{N_{\mathcal{S}_j}}{N_{\mathcal{S}}}
\]  

(42)

where \( N_{\mathcal{S}_j} \) is the number samples in \( \mathcal{S}_j \) for \( j = 1, 2 \), and \( N_{\mathcal{S}} \) is the number samples in \( \mathcal{S} \).

The variation at a node is defined as follows,

\[
g(\mathcal{S}_j) = \sum_{i=1}^{k} P(c_i|\mathcal{S}_j)(1 - P(c_i|\mathcal{S}_j))
\]

(43)

where \( g(\mathcal{S}_j) \) is maximum when the samples in the set \( \mathcal{S}_j \) are equally distributed into all possible classes and is minimum when the sample are all distributed into one class.

Using \( g(\mathcal{S}_j) \) the Gini Index is defined as
\[ G = P(S_1)g(S_1) + P(S_2)g(S_2) \]  \hspace{1cm} (44)

The feature and the threshold chosen at every node is the one that leads to lowest Gini Index in the above equation.

3.2.3.1.3 Random forests as ensemble decision trees [105]

Random forests are a collection of decision trees built via random selection of features at each node and sub-set of training data for each decision tree.

If \( h_k(x) \) is a single decision tree, then \( \mathcal{H} = \{ h_1(x), h_2(x), \ldots, h_k(x) \} \) represents the ensemble of decision trees.

If \( \mathcal{X} \) represents the data set that is used to train the random forest, each \( h_k(x) \) is trained using a data sub-set \( \mathcal{X}_k \subset \mathcal{X} \). For the random forest classifier, the feature used at each node of \( h_k(x) \) is obtained for a randomly chosen feature sub-set of all features.

Two classes were used in this study, i.e., spindles and non-spindles along with three features described in 3.2.3.4. The number of trees used was 10 and the number of features sampled at each node was 2.

3.2.3.2 Data

Only the MASS database was used in this particular study because the other databases contained no more than six subjects data in each. Due to the low number of subjects in the other databases, testing using a trained classifier would imply a very small dataset for testing and the results would not be reliable.

As previously described in the database section, the MASS database consists of data from 19 subjects, including 8 male and 11 female. The age of participants ranged between 18 and 33. All EEG channels were sampled at 256 Hz. Visual scoring was performed by two different scorers identified as V4 and V5. While V4 used AASM rules for scoring, V5 used a broader definition and a wider band (11-17Hz) to score SS. Scorings from V5 were only used to compare results in this study. Data from subjects 4, 8, 15 and 16 were excluded from analysis as sleep spindle scores from scorer V5 do not exist for these subjects.

3.2.3.3 Preprocessing to remove artifacts

The same procedure described in 3.1.2.2.1 was used to remove artifacts in this section.

3.2.3.4 Features used

Three features were generated as inputs to the Random Forest classifier. A moving window of 0.5 second length was used to develop all three features. The length of 0.5 seconds was chosen as this is
the minimal required length of a sleep spindle. IIR filters were used to extract frequency specific information.

\[ B_1(t) : S(t) \text{ filtered in the 12.5-15Hz band}; \]
\[ B_2(t) : S(t) \text{ filtered in the 10.5-16Hz band}; \]
\[ B_3(t) : S(t) \text{ filtered in the 4-10Hz band}; \]
\[ B_4(t) : S(t) \text{ filtered in the 20-40Hz band}; \]
\[ B_5(t) : S(t) \text{ filtered in the 8-12Hz band}; \]

1) **Alpha Ratio**
This is the ratio of the RMS amplitude in \( B_5(t) \) compared with the total RMS amplitude of the original signal. This feature provides information about the relative alpha activity in a 0.5 second moving window.

2) **Spindle Band Ratio**
The spindle band ratio is defined as the ratio of the RMS amplitude in \( B_2(t) \) to the total RMS amplitude of the original signal. This feature provides information about relative spindle activity in a 0.5 second moving window.

3) **Mean Sigma Index**
This is a modification version of a feature developed by Hupponean et al., and previously described in 3.2.2.2.3

\[ M_\sigma(t) = \frac{\mu(|B_1(t)|)}{\mu(|B_3(t)|) + \mu(|B_4(t)|)} \]

3.2.3.5 **Random forest training and testing**
To avoid over fitting, only 3 subjects were used for training and 12 subjects were used for testing purposes. Considering the large size of the training datasets and random forests classification being implemented for SS detection for the first time (0.5 second moving window through a whole night recording), the number of trees was set to 10 while the number of features to sample at each node was set to 2.

3.2.3.6 **Smoothing and removal of false detections**
A spindle was identified as a true spindle if more than 50% of the 0.5 second moving window samples were identified as spindles (others were rejected as false spindles). When this occurs, the whole 0.5 second window is marked as a true spindle.
3.3 Automated K-complex detection

3.3.1 Databases

Only one public database with visual scoring from two scorers was available to perform K-complex detection.

3.3.1.1 Dreams K-complex database

The Dreams® K-complex database is available online and was used to produce results. This database consists of 30-minute excerpts from 10 subjects. However, the visual scoring by two experts was only carried out on 5 subjects and these five subjects were used in this study. The results obtained from these first five subjects will be directly comparable with other studies that have also used only these same first five subjects. The PSG data consisted of three EEG channels all sampled at 200Hz. The five subjects used consisted of 4 female and 1 male subjects with age range 20-47 years.

3.3.2 Clustering method

The EM algorithm described in 3.1.2.3 was used for clustering the features.

3.3.2.1 EMG artifacts were removed using the same procedure described in 3.1.2.2.1

Features used –

Slope – this feature calculates the slope of the K-complex peak to peak wave within the moving window, and the slope of the K-complex, \( G_k \), is defined as

\[
G_k = \frac{p_+ - p_-}{T_{\text{max}} - T_{\text{min}}} \quad (45)
\]

where \( T_{\text{max}} \) is the time location of the positive peak in the moving window, \( T_{\text{min}} \) is the time location of the negative peak, \( p_+ \) is the positive peak amplitude and \( p_- \) is the negative peak amplitude.

Low delta index – \( I_{\Delta L} \) is defined as the power of the signal in the 0.5 Hz to 2 Hz range calculated from the STFT as follows,

\[
I_{\Delta L} = \int_{0.5}^{2} X_{\text{STFT}} (F, t) dF \quad (46)
\]

where \( X_{\text{STFT}} (F, t) dF \) represents the STFT of EEG signal \( x(t) \), \( F \) denotes frequency variable.

Results were obtained using different window sizes. Unlike sleep spindles where a refractory period prevents the occurrence of two consecutive sleep spindles occurring within 2-3 seconds of each other, K-complexes are known to occur sometimes in pair, one after the other. Since K-complex can occur in pairs, features similar to sigma ratio 2 used to compare a spindle to its immediate neighbourhood cannot be developed for the K-complex.
3.3.2.1 Removal of slow oscillations

Slow oscillations occurring in the NREM-3 stage of sleep and K-complexes are transient events that are very similar in signal structure [4]. They can be differentiated by the fact that slow oscillations consist of a sequence or train of K-complex like oscillations while K-complexes are isolated transients. Slow oscillations were filtered by checking if there were more than 3 K-complexes detected within a 20-second time window since more than 3 K-complexes in 20 seconds (20% of epoch) indicate NREM stage 3 [4]. After removal of slow oscillations, the remaining detections were all classified as true K-complexes.

3.3.3 Detection using pattern matched wavelets

3.3.3.1 Pattern matched wavelet development using CWT [107]

The continuous wavelet transform (CWT) is a signal analysis tool which uses a time-window function varying with time. It is defined as [107]

\[
X_{CWT}(t, s) = \int_{-\infty}^{\infty} x(\tau) \psi_{t,s}(\tau) d\tau
\]

(47)

where the signal \(x(\tau)\) is being decomposed using the wavelet function \(\psi_{t,s}(\tau)\) which is defined as

\[
\psi_{t,s}(\tau) = \frac{1}{\sqrt{s}} \psi\left(\frac{\tau - t}{s}\right)
\]

(48)

and \(s\) is the scaling factor \((s > 0)\) responsible for dilation and compression of the mother wavelet.

Pattern matched wavelet is a mother wavelet which has been designed to match a signal pattern of interest. A polynomial based method developed by Misisi et al. [107] was used to develop a pattern matched wavelet for K-complex detection. A single K-complex detected by a visual scorer from excerpt 1 in the database was used to design the wavelet. In order to develop a wavelet for use with the continuous wavelet transform, the wavelet needs to be admissible under the admissibility constraint defined as [107]

\[
\int_{-\infty}^{\infty} \frac{\left|\psi(\Omega)\right|^2}{|\Omega|} d\Omega < \infty
\]

(49)

where \(\psi(\Omega)\) is the Fourier transform of \(\psi(t)\). For a compactly supported wavelet for use with the CWT the admissibility condition can be reduced to the following equation [107]

\[
\int_{a}^{b} \psi(t) dt = 0
\]

(50)

The wavelet in the above equation is a zero mean function and has support in the interval \([a, b]\).
According to [107], a wavelet matching a specified pattern and satisfying (50) can be constructed using a family of functions $\mathcal{F} = \{\rho_i \mid 1 \leq i \leq N \}$. In this dissertation, polynomials of time up to degree 10 were used to construct the wavelet reducing (50) to the following,

$$\sum_{i=1}^{N} \alpha_i m_i = 0$$

(51)

where

$$m_i = \int_{a}^{b} \rho_i(t) dt$$

(52)

the unknown's $\alpha_i$ in the discrete sampled case can be solved using Lagrangian multipliers. The solution leads to the following equation [107]

$$\begin{pmatrix} G & m \end{pmatrix} \begin{pmatrix} \alpha \end{pmatrix} = \begin{pmatrix} b \end{pmatrix}$$

(53)

where $\lambda$ is the Lagrange multiplier and elements of matrix $G$, vectors $b$, $m$, $\alpha$ are given by

$$g_{ij} = \sum_{k=1}^{K} \rho_i(t_k) \rho_j(t_k),$$

(54)

$$m = [m_1, m_2, ..., m_N]^T,$$

(55)

$$\alpha = [\alpha_1, \alpha_2, ..., \alpha_N]^T,$$

(56)

$$b_i = \sum_{k=1}^{K} y_k \rho_i(t_k),$$

(57)

and $y_k$ is the original signal pattern that was matched to produce a wavelet.

A single $K$-complex from excerpt 1 identified by visual scorer 1 was initially used to produce the first matched wavelet. These are shown in Fig. 3-4. Use of $K$-complex derived from the mean of multiple visually scored $K$-complexes was avoided as this would result in a very smooth pattern. $K$-complexes by definition are defined as vertex sharp waves.
Figure 3-4 Pattern matched wavelet developed using Eq. 45. The process of matching is described in 3.3.3.1 along with the original K-complex used for matching.

An alternate second K-complex was also used to produce results, this allowed for a comparison of change in results based on initial choice of K-complex. The second K-complex and it’s matched wavelet are shown in Figure 3-5.

Figure 3-5. Pattern matched wavelet #2 developed using Eq. 45.
3.3.3.2 Continuous wavelet transform

Using the pattern matched wavelet from the previous section, CWT (as defined in (39)) was used to decompose the signals into multiple scales ranging from 0.02 seconds to 5.12 seconds with increments of 0.005 seconds. The C7-A1 channel was used for analysis of all subjects. This produced a CWT with very fine scale resolution. Scales higher than 5.12 seconds were not deemed useful as K-complexes are limited to a maximum duration of 2 seconds [4] and scales higher than 5.12 seconds contain very low frequency values which are not useful in the analysis. The wavelet transform along with the scales is shown in Fig. 3-6 for a 20-second sample from subject 1 of the Dreams K-complex database. The sample has a discernible K-complex in the given window.

![Scalogram and 20 second EEG signal with K-complex](image)

**Figure 3-6** A 20-second EEG segment with the wavelet scalogram. High coefficients for the scales around 1-second can be seen in correlation with the K-complex.

3.3.3.3 Detection of K-complexes

Since K-complexes are limited from 0.5 to 2 second duration, the maximum CWT coefficient for a K-complex should range within the scales of 0.5 seconds to 2 seconds. For each sample of time, the maximum coefficient from all scales was used to check if it was within the range of 0.5 seconds to 2 seconds. If the maximum of coefficients was within this range and the coefficient value was greater than the threshold, the sample was marked as a K-complex for further analysis. Minimum peak to peak amplitude for the K-complex was set to 80μV. Various thresholds ranging from 100 to 800 were used to assess results on database 1 (Fig. 4). A threshold of 400 was found to produce best results. Any artifact was automatically ignored as the maximum coefficient for artifacts are in the lower scales representing higher frequencies.
3.3.3.4 Removal of slow oscillations

The same procedure described in 3.3.2.1 was used to remove slow oscillations here.
4 Automated sleep transient detection techniques

4.1 Sleep spindle detection using STFT

The STFT method for sleep spindle detection described in 3.2.2 was evaluated for the Dreams database, MASS database and the Charité database. The results of the STFT method and the IIR filter method (section 4.2) were analysed to answer hypothesis 1 (sub-section 1.4.2) addressing the issue of automatic SS detection using clustering methods. For each database, results were obtained for varying moving window sizes. Furthermore, results were also obtained for each moving window size with different length checks to remove false spindles as described in sub-section 3.2.2.4. Detailed results are presented and analysed in the following sub-sections and in Chapter 7.

4.1.1 Dreams Database

Results were obtained for different window sizes and different SS length checks to remove false spindles. As described in sub-section 3.2.2.4, a length check implies that automatically detected sleep spindles below a certain length are not deemed to be spindles. Length was checked as percentage of moving window size. Although the AASM standard specifies a minimum length of 0.5 seconds, this constraint is not directly applicable here since the precise temporal localisation of a spindle within the moving window cannot be observed.

A receiver operator characteristic (ROC) curve is a plot of two variable statistics which change with change in a parameter used in an algorithm or a test. Traditional ROC curves which plot sensitivity vs (1-specificity) provide an inadequate representation of the optimisation problem in the case of SS due to high imbalance between the number of true positives and true negatives (SS are transient and only occur in NREM-2 and NREM-3 making the ratio of true positives to true negatives small) [112][42][26]. ROC like curves plotting sensitivity vs FPP or sensitivity vs FDR are better suited for the SS detection problem and were previous used in [26][42]. In this study the sensitivity vs FPP plot was used to compare results between different window sizes and different methods. All ROC curves presented here on are sensitivity vs FPP curves.

Fig. 4-1 shows the ROC curves for different moving window sizes, each curve’s variation is obtained by varying the minimum length check used to remove false spindles (going from 50% of window length to 120% of window length). Results closest to the top left hand corner of the plot, i.e., the point representing 100% sensitivity and 0% FPP represent the best trade-off between sensitivity and FPP [26][95]. In most typical ROC curves where an direct relationship between the sensitivity and FPP exists, the line intersecting with the diagonal connecting 100% Sensitivity and 100% FPP (shown as dashed line in Fig. 4.1) approximately produces the best trade-off between Sensitivity and FPP [26][95]. Ideal results were obtained using the 1.2 second window and a minimum length check of 1.2 seconds (100% of moving window size). Since ideal results are towards the top left hand corner of the plot, ROC curves leaning towards the top left hand corner also contain highest area under their curves. The Area Under Curve (AUC) statistic is also shown in Fig. 4-1. Although best results were
seen using the 1.2 second window, the AUC was minimally higher using the 1.35 second moving window (0.7535 vs 0.750).

![ROC curve for different length checks to remove false spindles](image)

Figure 4-1 ROC curves for different moving window sizes (Dreams database), the intersection of the curves with the dotted line show the approximate ideal trade-off between Sensitivity and False Positive Proportion. The legend describes the moving window length used for each curve, along with the area under the curve (AUC) for each curve. Curves with higher AUC lean towards the ideal results towards the top left hand corner of the plot.

As described in 3.2.2.4, the curves in Fig.4-1 were obtained by varying the minimum length check to remove false spindles. Ideal results for all moving window sizes were obtained using a 100% of moving window length check.

Fig. 4-2 shows the ROC curve for single length check (100% of window length) with different window sizes, i.e., false spindles for different window sizes were removed based on a length equal to the window size. Once again, results closest to the top left hand corner of the plot indicate best results. In this case, the 1.2 second window showed the best results.
Table 4-1 presents ideal results for the Dreams database using a 1.2 second window corresponding to the 1.2 second curve seen in Fig. 4-1. An overall sensitivity of 70.26% at an FPP of 28.25% was observed. Sensitivity for subjects ranged from 63.24% to 88.63%, while the FPP ranged from 12.82% to 75%. The sensitivity and FPP values obtained here are used to compare with the IIR filter method and existing literature in the discussion of Chapter 7.
Table 4-1 Ideal results obtained using a 1.2 second window, subjects are numbered S1, S2,…, S6. The Dreams database with two visual scorings is limited to six subjects on which a few existing algorithms have tested their performance.

<table>
<thead>
<tr>
<th>Subjects</th>
<th>Sensitivity %</th>
<th>False Positive Percentage %</th>
<th>Total number of spindles scored by visual scorers</th>
<th>True Positives</th>
<th>False Positives</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>67.16</td>
<td>14.17</td>
<td>134</td>
<td>90</td>
<td>19</td>
</tr>
<tr>
<td>S2</td>
<td>80.51</td>
<td>27.27</td>
<td>77</td>
<td>62</td>
<td>21</td>
</tr>
<tr>
<td>S3</td>
<td>88.63</td>
<td>75</td>
<td>44</td>
<td>39</td>
<td>33</td>
</tr>
<tr>
<td>S4</td>
<td>68.25</td>
<td>61.90</td>
<td>63</td>
<td>43</td>
<td>39</td>
</tr>
<tr>
<td>S5</td>
<td>67.96</td>
<td>24.27</td>
<td>103</td>
<td>70</td>
<td>25</td>
</tr>
<tr>
<td>S6</td>
<td>63.24</td>
<td>12.82</td>
<td>117</td>
<td>74</td>
<td>15</td>
</tr>
<tr>
<td>All Subjects</td>
<td>70.26</td>
<td>28.25</td>
<td>538</td>
<td>378</td>
<td>152</td>
</tr>
</tbody>
</table>

4.1.2 MASS database

Similar to the analysis conducted on the Dreams database, the results for the MASS database obtained for different window lengths and different minimum length checks to remove false spindles are presented here. A comparison between databases and between the STFT method and IIR filter method for each database is detailed in Chapter 7.

Fig. 4-3 shows the ROC curves for different moving window sizes, each curve’s variation is obtained by varying the minimum length check used to remove false spindles (going from 50% of window length to 120% of window length). Best results (closest to top left hand corner of plot) were obtained from the 0.8 second window at a false length check of 130% of window length. AUC was also highest for the 0.8 second window. Following the 0.8 second window, the 1 second window showed the highest AUC and best results at a 120% window length check. The difference in AUC between the 0.8 second window and the 1 second window was small (0.62049 vs 0.61114).
Figure 4-3 ROC curves for different moving window sizes (MASS database), the intersection of the curves with the dotted line show the approximate ideal trade-off between Sensitivity and False Positive Proportion.

Table 4-2 shows the ideal results obtained using a 1 second window and a 120% window length check corresponding to the 1 second ROC curve in Fig. 4-3. An overall sensitivity of 62.6% at an FPP of 38% was observed. Corresponding ROC curves are shown in Fig. 4-3. Individual subject sensitivities ranged from 47.3% to 73%, while FPP ranged from 15.6% to 62.8%. These results are later discussed (in Chapter 7) in comparison to existing literature and the IIR filters method.
Table 4-2 Ideal results obtained using a 1 second window for the MASS database. Subjects are numbered S1, S2,…, S15.

<table>
<thead>
<tr>
<th>Subjects</th>
<th>Sensitivity %</th>
<th>False Positive Percentage %</th>
<th>Total number of spindles scored by visual scorers</th>
<th>True Positives</th>
<th>False Positives</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>60.4</td>
<td>16.5</td>
<td>2448</td>
<td>1479</td>
<td>405</td>
</tr>
<tr>
<td>S2</td>
<td>64.8</td>
<td>15.6</td>
<td>2224</td>
<td>1441</td>
<td>348</td>
</tr>
<tr>
<td>S3</td>
<td>56.7</td>
<td>41.4</td>
<td>609</td>
<td>345</td>
<td>252</td>
</tr>
<tr>
<td>S4</td>
<td>47.3</td>
<td>30.8</td>
<td>1204</td>
<td>570</td>
<td>371</td>
</tr>
<tr>
<td>S5</td>
<td>52.9</td>
<td>46.0</td>
<td>847</td>
<td>448</td>
<td>390</td>
</tr>
<tr>
<td>S6</td>
<td>66.6</td>
<td>47.5</td>
<td>1672</td>
<td>1114</td>
<td>794</td>
</tr>
<tr>
<td>S7</td>
<td>73.0</td>
<td>62.8</td>
<td>1686</td>
<td>1231</td>
<td>1059</td>
</tr>
<tr>
<td>S8</td>
<td>50.4</td>
<td>18.5</td>
<td>1953</td>
<td>985</td>
<td>361</td>
</tr>
<tr>
<td>S9</td>
<td>71.2</td>
<td>46.1</td>
<td>1546</td>
<td>1100</td>
<td>713</td>
</tr>
<tr>
<td>S10</td>
<td>62.8</td>
<td>48.1</td>
<td>1246</td>
<td>782</td>
<td>599</td>
</tr>
<tr>
<td>S11</td>
<td>76.9</td>
<td>76.7</td>
<td>1465</td>
<td>1126</td>
<td>1123</td>
</tr>
<tr>
<td>S12</td>
<td>67.1</td>
<td>41.9</td>
<td>1639</td>
<td>1099</td>
<td>686</td>
</tr>
<tr>
<td>S13</td>
<td>71.8</td>
<td>49.1</td>
<td>1206</td>
<td>866</td>
<td>592</td>
</tr>
<tr>
<td>S14</td>
<td>46.8</td>
<td>13.5</td>
<td>1789</td>
<td>837</td>
<td>241</td>
</tr>
<tr>
<td>S15</td>
<td>67.2</td>
<td>61.1</td>
<td>1061</td>
<td>713</td>
<td>648</td>
</tr>
<tr>
<td>All subjects</td>
<td>62.6</td>
<td>38.0</td>
<td>22595</td>
<td>14136</td>
<td>8582</td>
</tr>
</tbody>
</table>

4.1.3 Charité Database

Similar to the results obtained on the Dreams database and the MASS database, Fig. 4-4 shows the results obtained for different moving window sizes and by varying the minimum length checks to remove false spindles. Best results were obtained using the 1 second moving window with a length check of 120% of window length. An overall sensitivity of 44.81% at an FPP of 58.2% was recorded. The 1 second window also showed the highest AUC of 0.36. Unlike the sensitivity and FPP measures, the F₁-score statistic described in section 3.1 is a single measure that allows direct comparison between results for different databases. As described in section 3.1, it is the harmonic mean of the sensitivity and precision statistics. An F₁-score of 1 is synonymous with correct classification of all SS and non-spindles and an F₁-score of 0 implies no SS or non-spindles were classified correctly [16]. Results for the Charité database showed poorer results in terms of F₁-score compared to the other two databases. While the F₁-scores of ideal results for the Dreams database and the MASS database were 0.70 and 0.62 respectively, the F₁-score of ideal results for the Charité Database was 0.44.

Table 4-3 shows the results obtained using the 1 second moving window.
Table 4-3 Ideal results obtained using a 1 second window for the Charité database, subjects are numbered S1, S2,..., S6

<table>
<thead>
<tr>
<th>Subjects</th>
<th>Sensitivity %</th>
<th>False Positive Percentage %</th>
<th>Total number of spindles scored by visual scorers</th>
<th>True Positives</th>
<th>False Positives</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>46.53</td>
<td>59.6</td>
<td>750</td>
<td>349</td>
<td>447</td>
</tr>
<tr>
<td>S2</td>
<td>66.75</td>
<td>69.11</td>
<td>761</td>
<td>508</td>
<td>526</td>
</tr>
<tr>
<td>S3</td>
<td>25</td>
<td>30.25</td>
<td>780</td>
<td>195</td>
<td>236</td>
</tr>
<tr>
<td>S4</td>
<td>37.57</td>
<td>83.46</td>
<td>889</td>
<td>334</td>
<td>742</td>
</tr>
<tr>
<td>S5</td>
<td>73.99</td>
<td>115.08</td>
<td>769</td>
<td>569</td>
<td>885</td>
</tr>
<tr>
<td>S6</td>
<td>31.39</td>
<td>19.216</td>
<td>1379</td>
<td>433</td>
<td>265</td>
</tr>
<tr>
<td><strong>All Subjects</strong></td>
<td><strong>44.81</strong></td>
<td><strong>58.20</strong></td>
<td><strong>5328</strong></td>
<td><strong>2388</strong></td>
<td><strong>3101</strong></td>
</tr>
</tbody>
</table>
Figure 4-4 ROC curves for different moving window sizes (Charité Database), the intersection of the curves with the dotted line show the approximate ideal trade-off between Sensitivity and False Positive Proportion.

4.2 Sleep spindle detection using IIR filters.

Performance evaluation of STFT method for sleep spindle detection using different ideal window size parameters for different databases indicated over fitting of parameters to each database and, therefore, a universal 1 second window was used to evaluate performance of the IIR filtering method for sleep spindle detection.

The IIR filter was also computationally faster. Fig. 4-5 shows a comparison of processing time required to obtain results for a single 8 hour sleep EEG recording using both the STFT method and the IIR filters method executed on a MATLAB 2015b platform running on an Intel i7-core CPU @ 3.6 Ghz.
Figure 4-5 Comparison of processing time required by the STFT based clustering method and the IIR filters based clustering method, executed on a MATLAB 2015b platform running on an Intel i7-core CPU @ 3.6 Ghz. Processing time for event detection was significantly reduced using IIR filters.

Fig. 4-6 shows the ROC curves for the STFT method and the IIR method using a 1 second window and different length checks to remove false spindles (using the MASS database since it provided the most number of subjects). This allows a direct comparison of the two methods. Ideal results using the IIR filters method was higher along with a higher AUC.
A comparison between results obtained by the optimised STFT method (1 second moving window, 120% length check) and those by the IIR method (1 second moving window, 90% length check) using $F_1$-scores is shown in Table 4-4. As described in section 3.1, the $F_1$-score is the harmonic mean of sensitivity and precision statistics. A high $F_1$-score indicates better performance. The overall $F_1$-score for the IIR method was higher but not significant when a paired t-test (Appendix B) was performed ($p > 0.05$). It should be noted that IIR method was not optimised for each database to avoid over fitting. The signal from the C3 channel was used to evaluate both methods. The C3 channel was the channel used for visual scoring of the MASS database.
Table 4-4 Comparison between the STFT method and the IIR filters method using the F₁-score statistic. The IIR method showed better performance in terms of producing a higher F₁-statistic.

<table>
<thead>
<tr>
<th>Subjects</th>
<th>F₁-score STFT method</th>
<th>F₁-score IIR method</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>0.68</td>
<td>0.71</td>
</tr>
<tr>
<td>S2</td>
<td>0.72</td>
<td>0.72</td>
</tr>
<tr>
<td>S3</td>
<td>0.57</td>
<td>0.51</td>
</tr>
<tr>
<td>S4</td>
<td>0.53</td>
<td>0.57</td>
</tr>
<tr>
<td>S5</td>
<td>0.53</td>
<td>0.51</td>
</tr>
<tr>
<td>S6</td>
<td>0.62</td>
<td>0.65</td>
</tr>
<tr>
<td>S7</td>
<td>0.62</td>
<td>0.71</td>
</tr>
<tr>
<td>S8</td>
<td>0.60</td>
<td>0.68</td>
</tr>
<tr>
<td>S9</td>
<td>0.65</td>
<td>0.66</td>
</tr>
<tr>
<td>S10</td>
<td>0.60</td>
<td>0.60</td>
</tr>
<tr>
<td>S11</td>
<td>0.61</td>
<td>0.54</td>
</tr>
<tr>
<td>S12</td>
<td>0.64</td>
<td>0.59</td>
</tr>
<tr>
<td>S13</td>
<td>0.65</td>
<td>0.61</td>
</tr>
<tr>
<td>S14</td>
<td>0.58</td>
<td>0.67</td>
</tr>
<tr>
<td>S15</td>
<td>0.59</td>
<td>0.60</td>
</tr>
<tr>
<td>All subjects</td>
<td>0.6239</td>
<td>0.6343</td>
</tr>
</tbody>
</table>

4.2.1 Dreams Database

Results for subject 3 using the IIR method were not possible due to the low sampling frequency of 50 Hz; this was not a problem with the STFT method in the earlier section as frequency features were restricted to the SS frequency range (<16Hz). The mean sigma index feature in the IIR filters method (described in sub-section 3.2.2.2.3) require filtering of the montage signal in the 20-40 Hz frequency range. A 50 Hz sampling frequency implies a Nyquist frequency of 25 Hz, therefore the mean sigma index feature cannot be calculated for this subject and hence the subject was excluded.

Table 4-5 IIR filters method results obtained using a 1 second moving window, subjects are numbered S1, S2, ..., S6. Subject 3 has been excluded due to low sampling rate which did not allow the calculation of the necessary mean sigma index feature. The Dreams database with two visual scorings is limited to six subjects on which a few existing algorithms have tested their performance.

<table>
<thead>
<tr>
<th>Subjects</th>
<th>Sensitivity %</th>
<th>False Positive Percentage %</th>
<th>Total number of spindles scored by visual scorers</th>
<th>True Positives</th>
<th>False Positives</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>77.61</td>
<td>29.10</td>
<td>134</td>
<td>104</td>
<td>39</td>
</tr>
<tr>
<td>S2</td>
<td>87.01</td>
<td>106.49</td>
<td>77</td>
<td>67</td>
<td>82</td>
</tr>
<tr>
<td>S4</td>
<td>100.00</td>
<td>195.24</td>
<td>63</td>
<td>63</td>
<td>123</td>
</tr>
<tr>
<td>S5</td>
<td>69.90</td>
<td>39.81</td>
<td>103</td>
<td>72</td>
<td>41</td>
</tr>
<tr>
<td>S6</td>
<td>62.39</td>
<td>16.24</td>
<td>117</td>
<td>73</td>
<td>19</td>
</tr>
<tr>
<td>All Subjects</td>
<td>76.72</td>
<td>61.54</td>
<td>494</td>
<td>379</td>
<td>304</td>
</tr>
</tbody>
</table>
Table 4-5 shows the results using the IIR method (with 1 second moving window) for sleep spindle detection based on the Dreams database. An overall sensitivity of 76.7% at FPP of 61.5% was produced. Individual subject sensitivity ranged from 62.39% to 100% while the FPP ranged from 16.24% to 195.24%. A detailed comparison of these results with the STFT method and existing literature is discussed in Chapter 7. Using a similar 1 second window and a 90% window length check with the STFT method showed a sensitivity of 89% at an FPP of 62.8% i.e the STFT method showed better performance for this database.

![ROC curve using STFT and IIR filter methods (1 second moving window)](image)

Figure 4-7 Comparison of ROC curves of the STFT and IIR filter methods (Dreams database). ROC curves were obtained by varying length checks to remove false spindles using a 1 second moving window. In this case, the STFT method showed better performance in terms of higher AUC.
4.2.2 MASS database

The MASS database, being larger than Dream database allowed for a detailed analysis of results using different montages as described in sub-section 3.2.2.2.3. Results for sleep spindle detection using IIR filters were produced for montages and individual channels. Results were analysed for multiple factors to check for slow-spindle fast spindle dichotomy, presence of NREM-3 spindles, frequency distribution of spindles and correlation of automatic scored spindles to visually scored spindles. They are presented in the following tables and figures. Detailed comparison and interpretation of these results is provided in Chapter 7. A direct comparison between the STFT method and IIR filters method for this database showed better performance of the IIR filters method in terms of sensitivity, FPP and F₁-score. The Table 4-6 shows the results obtained using a 1 second moving window for the MASS database using a C₃ + C₂ + C₄ montage (i.e., Central channels). An overall sensitivity of 70%, F₁-score of 0.59 and a false positive proportion of 66% were obtained.

<table>
<thead>
<tr>
<th>Subjects</th>
<th>True Positives</th>
<th>False Positives</th>
<th>Total number of spindles scored by visual scorer</th>
<th>Sensitivity (%)</th>
<th>Precision</th>
<th>F1-score</th>
<th>False Positive proportion (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1635</td>
<td>782</td>
<td>2448</td>
<td>67</td>
<td>0.68</td>
<td>0.67</td>
<td>32</td>
</tr>
<tr>
<td>2</td>
<td>1617</td>
<td>1086</td>
<td>2224</td>
<td>73</td>
<td>0.60</td>
<td>0.66</td>
<td>49</td>
</tr>
<tr>
<td>3</td>
<td>315</td>
<td>367</td>
<td>609</td>
<td>52</td>
<td>0.46</td>
<td>0.49</td>
<td>60</td>
</tr>
<tr>
<td>5</td>
<td>729</td>
<td>851</td>
<td>1204</td>
<td>61</td>
<td>0.46</td>
<td>0.52</td>
<td>71</td>
</tr>
<tr>
<td>6</td>
<td>553</td>
<td>1014</td>
<td>847</td>
<td>65</td>
<td>0.35</td>
<td>0.46</td>
<td>120</td>
</tr>
<tr>
<td>7</td>
<td>1199</td>
<td>1024</td>
<td>1672</td>
<td>72</td>
<td>0.54</td>
<td>0.62</td>
<td>61</td>
</tr>
<tr>
<td>9</td>
<td>1147</td>
<td>575</td>
<td>1686</td>
<td>68</td>
<td>0.67</td>
<td>0.67</td>
<td>34</td>
</tr>
<tr>
<td>10</td>
<td>1189</td>
<td>727</td>
<td>1953</td>
<td>61</td>
<td>0.62</td>
<td>0.61</td>
<td>37</td>
</tr>
<tr>
<td>11</td>
<td>1187</td>
<td>995</td>
<td>1546</td>
<td>77</td>
<td>0.54</td>
<td>0.64</td>
<td>64</td>
</tr>
<tr>
<td>12</td>
<td>859</td>
<td>1021</td>
<td>1246</td>
<td>69</td>
<td>0.46</td>
<td>0.55</td>
<td>82</td>
</tr>
<tr>
<td>13</td>
<td>1145</td>
<td>1956</td>
<td>1465</td>
<td>78</td>
<td>0.37</td>
<td>0.50</td>
<td>134</td>
</tr>
<tr>
<td>14</td>
<td>1137</td>
<td>1117</td>
<td>1639</td>
<td>69</td>
<td>0.50</td>
<td>0.58</td>
<td>68</td>
</tr>
<tr>
<td>17</td>
<td>978</td>
<td>1315</td>
<td>1206</td>
<td>81</td>
<td>0.43</td>
<td>0.56</td>
<td>109</td>
</tr>
<tr>
<td>18</td>
<td>1249</td>
<td>912</td>
<td>1789</td>
<td>70</td>
<td>0.58</td>
<td>0.63</td>
<td>51</td>
</tr>
<tr>
<td>19</td>
<td>840</td>
<td>1122</td>
<td>1061</td>
<td>79</td>
<td>0.43</td>
<td>0.56</td>
<td>106</td>
</tr>
<tr>
<td>All</td>
<td>15779</td>
<td>14864</td>
<td>22595</td>
<td>70</td>
<td>0.51</td>
<td>0.59</td>
<td>66</td>
</tr>
</tbody>
</table>
As described in sub-section 3.2.2.3, results were obtained for different montages (using averaged signal from multiple electrodes and single electrodes). Table 4-7 describes the results obtained using the different montages. The best results in terms of the F1-score measure were obtained using the C3 signal. The C3 signal was used for visual scoring, this is possible explanation of better performance on the C3 channel. However an ANOVA test (Appendix B) to compare these results between different montages did not show any significant differences between the C3 channels and other montages ($p > 0.05$). Fig 4-8 shows results comparison of these montages.

<table>
<thead>
<tr>
<th>Montage</th>
<th>Sensitivity %</th>
<th>FPP %</th>
<th>Precision</th>
<th>F1-score</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>C3+Cz+C4</td>
<td>69.40</td>
<td>7.88</td>
<td>71.85</td>
<td>31.9</td>
</tr>
<tr>
<td>F3+F4+C3+</td>
<td>69.19</td>
<td>7.89</td>
<td>71.07</td>
<td>31.2</td>
</tr>
<tr>
<td>Cz+C4+P3+</td>
<td>Cz+P4</td>
<td>69.95</td>
<td>8.91</td>
<td>69.42</td>
</tr>
<tr>
<td>C3+Cz+C4+</td>
<td>P3+Pz+P4</td>
<td>67.72</td>
<td>7.78</td>
<td>72.98</td>
</tr>
<tr>
<td>F3</td>
<td>68.95</td>
<td>9.32</td>
<td>73.34</td>
<td>32.8</td>
</tr>
<tr>
<td>F4</td>
<td>59.65</td>
<td>8.84</td>
<td>74.08</td>
<td>35.0</td>
</tr>
<tr>
<td>C3</td>
<td>73.44</td>
<td>7.79</td>
<td>64.71</td>
<td>26.8</td>
</tr>
<tr>
<td>C4</td>
<td>67.64</td>
<td>7.26</td>
<td>74.75</td>
<td>35.3</td>
</tr>
<tr>
<td>Cz</td>
<td>61.90</td>
<td>11.09</td>
<td>73.15</td>
<td>28.6</td>
</tr>
<tr>
<td>P3</td>
<td>61.31</td>
<td>15.35</td>
<td>67.15</td>
<td>26.9</td>
</tr>
<tr>
<td>P4</td>
<td>57.57</td>
<td>11.78</td>
<td>70.97</td>
<td>32.1</td>
</tr>
</tbody>
</table>
Figure 4-8 Sensitivity, FPP, precision and F1-score for different montages of the Montreal Archive of Sleep Studies database (MASS) shown in Table 4-7. Highest sensitivity is seen using the C3 channel which also showed the lowest FPP, highest F1-score and highest precision. Poorest results are seen using the m4 montage which includes central and parietal derivations.

Slow and fast spindles were separated using the spectrum of each individual spindle by fast Fourier transform (FFT) algorithm. A zero padded 1024-point FFT was used to calculate the centre frequency of a spindle in the 10.5-16 Hz frequency band, and the frequency which showed maximum amplitude in the 10.5-16 Hz range was used to classify the spindles as either a slow spindle or fast spindle. A 13 Hz cut-off was used based on [113], and a 14 Hz cut-off based on [114]. Table 4-8 and Fig. 4-9 summarise these results. Fig. 4-10 shows the distribution of the centre frequency of spindles. The results in Table 4-8 and Fig. 4-9, 10 indicate higher number of slow spindles detected using the frontal derivations and higher number of fast spindles detected using the posterior distributions. This is discussed further in sub-section 7.1.1.3.3. No visually discernible slow spindle-fast spindle dichotomy is seen in Fig. 4-10, further discussions in sub-section 7.1.1.3.3.
Table 4-8 Mass Database Slow and Fast Spindle Results, two different cut-offs were used. 13 Hz cut-off (De Gennaro and Ferrara, 2003) and 14 Hz cut-off (Dang-Vu et al., 2003). A shift of majority is seen when the cut-off is changed from 13 Hz to 14 Hz. This implies a majority of the spindle scores were between the 13Hz to 14 Hz range. When comparing the frequency distributions of the spindles scored by the algorithm and the spindles scored by visual scorers, no significant differences were observed as the one-way ANOVA showed p>0.05.

<table>
<thead>
<tr>
<th>Montage</th>
<th>Slow Spindles</th>
<th></th>
<th>Fast Spindles</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>&lt;=13 Hz</td>
<td>&lt;=14 Hz</td>
<td>&gt;13 Hz</td>
<td>&gt; 14 Hz</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>C3+Cz+C4</td>
<td>594</td>
<td>298</td>
<td>1627</td>
<td>655</td>
</tr>
<tr>
<td></td>
<td>1460</td>
<td>400</td>
<td>427</td>
<td>325</td>
</tr>
<tr>
<td>F3+F4+C3+Cz+C4+P3+Pz+P4</td>
<td>594</td>
<td>277</td>
<td>1622</td>
<td>631</td>
</tr>
<tr>
<td></td>
<td>1447</td>
<td>410</td>
<td>418</td>
<td>333</td>
</tr>
<tr>
<td>F3+F4+C3+Cz+C4</td>
<td>664</td>
<td>321</td>
<td>1652</td>
<td>628</td>
</tr>
<tr>
<td></td>
<td>1364</td>
<td>396</td>
<td>376</td>
<td>298</td>
</tr>
<tr>
<td>C3+Cz+C4+P3+Pz+P4</td>
<td>570</td>
<td>266</td>
<td>1612</td>
<td>625</td>
</tr>
<tr>
<td></td>
<td>1475</td>
<td>403</td>
<td>433</td>
<td>332</td>
</tr>
<tr>
<td>F3</td>
<td>1039</td>
<td>532</td>
<td>1817</td>
<td>588</td>
</tr>
<tr>
<td></td>
<td>1036</td>
<td>351</td>
<td>258</td>
<td>207</td>
</tr>
<tr>
<td>F4</td>
<td>898</td>
<td>505</td>
<td>1655</td>
<td>512</td>
</tr>
<tr>
<td></td>
<td>1027</td>
<td>396</td>
<td>270</td>
<td>220</td>
</tr>
<tr>
<td>C3</td>
<td>593</td>
<td>266</td>
<td>1600</td>
<td>627</td>
</tr>
<tr>
<td></td>
<td>1431</td>
<td>418</td>
<td>424</td>
<td>328</td>
</tr>
<tr>
<td>C4</td>
<td>628</td>
<td>340</td>
<td>1638</td>
<td>624</td>
</tr>
<tr>
<td></td>
<td>1430</td>
<td>378</td>
<td>420</td>
<td>311</td>
</tr>
<tr>
<td>Cz</td>
<td>663</td>
<td>301</td>
<td>1585</td>
<td>574</td>
</tr>
<tr>
<td></td>
<td>1310</td>
<td>393</td>
<td>388</td>
<td>270</td>
</tr>
<tr>
<td>P3</td>
<td>568</td>
<td>225</td>
<td>1479</td>
<td>594</td>
</tr>
<tr>
<td></td>
<td>1292</td>
<td>448</td>
<td>381</td>
<td>286</td>
</tr>
<tr>
<td>P4</td>
<td>658</td>
<td>268</td>
<td>1497</td>
<td>538</td>
</tr>
<tr>
<td></td>
<td>1221</td>
<td>407</td>
<td>382</td>
<td>268</td>
</tr>
<tr>
<td>Visually scored spindles</td>
<td>279</td>
<td>128</td>
<td>1122</td>
<td>516</td>
</tr>
<tr>
<td></td>
<td>1227</td>
<td>447</td>
<td>352</td>
<td>339</td>
</tr>
</tbody>
</table>

Figure 4-9 Slow and Fast spindle counts for different montages. A higher number of slow spindles can be seen when using the frontal derivations in m2 and a higher number of fast spindles are seen in the parietal derivations; earlier studies examining the slow fast dichotomy showed similar differences between frontal slow spindles and parietal slow spindles [15][16].
Figure 4-10 Central frequency distribution of sleep spindles identified by automatic detection and visual scoring. No visually discernible slow-fast dichotomy can be seen in the plots including visual scoring. Slow spindles can be seen in the plots showing automatic detections using different montages, however very few slow spindles can be seen in the plot showing frequency distribution of the visually scored spindles, suggesting that most visual scored spindles were centro-parietal fast spindles.

The visual scorers for both the MASS database and the Charité database did not score spindles in the NREM-3 sleep stage. Table 4-9 shows results when the NREM-3 spindles are excluded and included. The exclusion of NREM-3 spindles shows lower FPP and higher F$_1$-scores. The higher F$_1$-score (section 3.1) implies better performance of the IIR filters method when NREM-3 spindles are excluded from automatic scoring or included in visual scoring. F$_1$-scores using the C3 signal showed improvements from 0.63 to 0.69 when NREM-3 spindles were excluded. Although results of the IIR filters method using the Charité database are presented in the next section, results are also included in Table 4-9 for comparison.
Table 4-9 Results showing differences in False Positive Proportion and F$_1$-score when NREM 3 spindles are not considered in the MASS database (D1) and Charité database (D2). Data from all subjects was combined to obtain the results in this table.

<table>
<thead>
<tr>
<th>Montage</th>
<th>Combined Sensitivity for all subjects %</th>
<th>Combined FPP for all subjects %</th>
<th>Combined F$_1$-score for all subjects</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>With NREM 3</td>
<td>Without NREM 3</td>
<td>With NREM 3</td>
</tr>
<tr>
<td>D1: C3+Cz+C4</td>
<td>69.8</td>
<td>69.8</td>
<td>65.8</td>
</tr>
<tr>
<td>D1: F3+F4+C3+Cz+C4 +P3+Pz+P4</td>
<td>69.6</td>
<td>69.6</td>
<td>64.7</td>
</tr>
<tr>
<td>D1: F3+F4+C3+Cz+C4</td>
<td>70.5</td>
<td>70.5</td>
<td>63.3</td>
</tr>
<tr>
<td>D1: C3+Cz+C4+P3+Pz +P4</td>
<td>68.2</td>
<td>68.2</td>
<td>66.4</td>
</tr>
<tr>
<td>D1: F3</td>
<td>69.6</td>
<td>69.6</td>
<td>67.0</td>
</tr>
<tr>
<td>D1: F4</td>
<td>60.6</td>
<td>60.6</td>
<td>66.2</td>
</tr>
<tr>
<td>D1: C3</td>
<td>74.1</td>
<td>74.1</td>
<td>59.6</td>
</tr>
<tr>
<td>D1: C4</td>
<td>67.8</td>
<td>67.8</td>
<td>67.5</td>
</tr>
<tr>
<td>D1: Cz</td>
<td>62.9</td>
<td>62.9</td>
<td>67.2</td>
</tr>
<tr>
<td>D1: P3</td>
<td>61.3</td>
<td>61.3</td>
<td>61.1</td>
</tr>
<tr>
<td>D1: P4</td>
<td>58.1</td>
<td>58.1</td>
<td>63.9</td>
</tr>
<tr>
<td>D2: C3+C4</td>
<td>65.18</td>
<td>65.18</td>
<td>119.74</td>
</tr>
</tbody>
</table>

All previous results presented used the union of scorings of two visual scorers to calculate the results. Results were also calculated using an intersection of visual scoring, i.e., using the visual scorings in which both scorers agree on a sleep spindle. These results shown in Table 4-10 are for basic comparison only. While the sensitivity using the union of scoring was 74.1%, the sensitivity using the intersection of scoring was 80.2% with a corresponding increase in FPP.
Table 4-10 Results comparing the union of scoring and the intersection of scoring for the MASS database using the C3 channel. Data from all subjects was combined to obtain the results in this table.

<table>
<thead>
<tr>
<th>Visual scoring type</th>
<th>Combined for all subjects %</th>
<th>Combined FPP for all subjects %</th>
<th>Combined F₁-score for all subjects</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>With NREM 3</td>
<td>Without NREM 3</td>
<td>With NREM 3</td>
</tr>
<tr>
<td>V4 U V5</td>
<td>74.1</td>
<td>74.1</td>
<td>59.6</td>
</tr>
<tr>
<td>V4 ∩ V5</td>
<td>80.2</td>
<td>80.2</td>
<td>233.7</td>
</tr>
</tbody>
</table>

Table 4-11 shows the Pearson correlation (Appendix B) between the total number of spindles scored by the visual scorers and the total number of spindles scored by IIR filters method. A higher Pearson correlation indicated that the automatic detection algorithm is able to capture the inter subject difference better than an algorithm with a lower Pearson correlation. The p-value (Appendix B) for all montages of the MASS database (represented as D1 in the table) showed significant results (P<0.05) indicating that the algorithm was able to capture the inter-subject differences in total number of spindles. The Charité database is represented as ‘D2’ in the table.

Table 4-11 Results showing the Pearson’s correlation coefficient between the number of spindles scored by the visual scorers versus the number of spindles scored by the automatic detection algorithm for different subjects.

<table>
<thead>
<tr>
<th>Montage</th>
<th>Pearson’s Correlation Coefficient (r)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>D1: C3+Cz+C4</td>
<td>0.63</td>
<td>0.0113</td>
</tr>
<tr>
<td>D1: F3+F4+C3+Cz+C4 +P3+Pz+P4</td>
<td>0.63</td>
<td>0.0115</td>
</tr>
<tr>
<td>D1: F3+F4+C3+Cz+C4</td>
<td>0.63</td>
<td>0.0114</td>
</tr>
<tr>
<td>D1: C3+Cz+C4+P3+Pz +P4</td>
<td>0.62</td>
<td>0.0123</td>
</tr>
<tr>
<td>D1: F3</td>
<td>0.55</td>
<td>0.0311</td>
</tr>
<tr>
<td>D1: F4</td>
<td>0.59</td>
<td>0.0192</td>
</tr>
<tr>
<td>D1: C3</td>
<td>0.71</td>
<td>0.0027</td>
</tr>
<tr>
<td>D1: C4</td>
<td>0.56</td>
<td>0.0293</td>
</tr>
<tr>
<td>D1: Cz</td>
<td>0.65</td>
<td>0.0085</td>
</tr>
<tr>
<td>D1: P3</td>
<td>0.52</td>
<td>0.0424</td>
</tr>
<tr>
<td>D1: P4</td>
<td>0.49</td>
<td>0.0585</td>
</tr>
<tr>
<td>D2: C3+C4</td>
<td>0.53</td>
<td>0.2804</td>
</tr>
</tbody>
</table>
Table 4-12 shows the distribution of spindles in different sleep stages scored by the IIR filters method devised in this work in comparison with another algorithm by Ray et al. [115] and the visual scorers. Spindle density of automatically scored spindles in NREM-2 was similar to visually scored spindles. Highest sleep density was seen using the F3 channel followed by the C3 channel. Highest number of false spindles detected in the awake stage (Spindles are not existent in the awake stage [4]) were seen using the P3 channel. Proximity of P3 channel to the occipital lobe and therefore a proximity to occipital alpha artifacts are potential causes [4].

Table 4-12 Results showing spindle density (spindles/minute) for different sleep stages. The mean (standard deviation) have been shown using our method, Visual scoring and Ray et al., [115] using the MASS database.

<table>
<thead>
<tr>
<th>Method</th>
<th>Awake</th>
<th>NREM1</th>
<th>NREM2</th>
<th>NREM3 + NREM4</th>
<th>REM</th>
</tr>
</thead>
<tbody>
<tr>
<td>IIR filters Method C3</td>
<td>0.55(0.38)</td>
<td>0.93(0.52)</td>
<td>6.12(1.49)</td>
<td>3.85(1.58)</td>
<td>0.29(0.15)</td>
</tr>
<tr>
<td>IIR filters Method P3</td>
<td>1.67(1.54)</td>
<td>0.83(0.54)</td>
<td>5.30(1.7)</td>
<td>3.27(0.95)</td>
<td>0.47(0.31)</td>
</tr>
<tr>
<td>IIR filters Method F3</td>
<td>0.59(0.33)</td>
<td>1.13(0.66)</td>
<td>6.18(1.39)</td>
<td>4.09(2.15)</td>
<td>0.37(0.29)</td>
</tr>
<tr>
<td>Union of Visual scoring (C3)</td>
<td></td>
<td></td>
<td></td>
<td>5.71(1.82)</td>
<td></td>
</tr>
<tr>
<td>Ray et al., 2015 P3</td>
<td>NA</td>
<td>NA</td>
<td>4.83(0.3)</td>
<td>5.15(0.37)</td>
<td>NA</td>
</tr>
<tr>
<td>Ray et al., 2015 F3</td>
<td>NA</td>
<td>NA</td>
<td>4.54(0.21)</td>
<td>5.25(0.25)</td>
<td>NA</td>
</tr>
</tbody>
</table>
4.2.3 Charité Database.

Table 4-13 contains the results obtained for the Charité database using the IIR filters method with a 1 second moving window. An overall Sensitivity of 65.18% was observed at a False Positive Proportion of 119.74%. Individual sensitivity ranges from 36.67% to 83.62%, and FPP ranges from 60.26% to 177.63%.

Table 4-13, Charité database results using the IIR filters method. An overall sensitivity of 65.18% at an FPP of 119.74% was observed.

<table>
<thead>
<tr>
<th>Subjects</th>
<th>True Positive</th>
<th>False Positives</th>
<th>Total number scored by visual scorer</th>
<th>Sensitivity (%)</th>
<th>Precision</th>
<th>F1-score</th>
<th>False Positive Proportion (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>491</td>
<td>1010</td>
<td>750</td>
<td>65.47</td>
<td>0.33</td>
<td>0.44</td>
<td>134.67</td>
</tr>
<tr>
<td>2</td>
<td>625</td>
<td>1121</td>
<td>761</td>
<td>82.13</td>
<td>0.36</td>
<td>0.50</td>
<td>147.31</td>
</tr>
<tr>
<td>3</td>
<td>286</td>
<td>470</td>
<td>780</td>
<td>36.67</td>
<td>0.38</td>
<td>0.37</td>
<td>60.26</td>
</tr>
<tr>
<td>4</td>
<td>509</td>
<td>1148</td>
<td>889</td>
<td>57.26</td>
<td>0.31</td>
<td>0.40</td>
<td>129.13</td>
</tr>
<tr>
<td>5</td>
<td>643</td>
<td>1366</td>
<td>769</td>
<td>83.62</td>
<td>0.32</td>
<td>0.46</td>
<td>177.63</td>
</tr>
<tr>
<td>6</td>
<td>919</td>
<td>1265</td>
<td>1379</td>
<td>66.64</td>
<td>0.42</td>
<td>0.52</td>
<td>91.73</td>
</tr>
<tr>
<td>All</td>
<td>3473</td>
<td>6380</td>
<td>5328</td>
<td>65.18</td>
<td>0.35</td>
<td>0.46</td>
<td>119.74</td>
</tr>
</tbody>
</table>

Similar to ROC curves obtained for the Dreams database and the MASS database, Fig. 4-5 shows the ROC curves for the STFT and IIR filter methods using a 1 second window while the length check was varied. The IIR filters method showed the higher AUC of 0.39296.
Figure 4-11 Comparison of ROC curves of the STFT and IIR filter methods (Charite database). ROC curves were obtained by varying minimum length checks (sub-section 3.2.2.4) to remove false spindles using a 1 second moving window.

4.2.4 Results using the random forest classifier—
Results for the MASS database were also obtained using the random forest classifier method described in section 3.2.3. Three whole night recordings from the MASS database were used to train the classifier and twelve whole night recordings were used in the testing set. Only scorings from visual scorer #2 were used as this provided a larger testing set. Given the small number of subjects in the Dreams database and the Charité database, the classifier was not trained on those sets. Table 4-14 shows the results of the testing set from the MASS database. An overall sensitivity of 71.2% was observed at an FPP of 64.27%. Sensitivity ranged from 46.9% to 84.2%, while FPP ranged from 27.42% to 138.7%. These results are discussed in comparison to other methods and existing literature in sub-section 7.1.1.2
Table 4-14 Results obtained using the Random Forrest Classifier (MASS database).

<table>
<thead>
<tr>
<th>Subjects</th>
<th>Sensitivity %</th>
<th>Specificity %</th>
<th>False Positive Proportion %</th>
<th># of Spindles scored by visual scorer</th>
<th>True Positives</th>
<th>False Positives</th>
<th>F₁-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>S5</td>
<td>58.3</td>
<td>98.5</td>
<td>37.56</td>
<td>1198</td>
<td>699</td>
<td>450</td>
<td>0.60</td>
</tr>
<tr>
<td>S6</td>
<td>46.9</td>
<td>99.0</td>
<td>31.54</td>
<td>837</td>
<td>393</td>
<td>264</td>
<td>0.53</td>
</tr>
<tr>
<td>S7</td>
<td>74</td>
<td>96.8</td>
<td>53.09</td>
<td>1601</td>
<td>1186</td>
<td>850</td>
<td>0.65</td>
</tr>
<tr>
<td>S9</td>
<td>66.3</td>
<td>98.3</td>
<td>27.42</td>
<td>1663</td>
<td>1103</td>
<td>456</td>
<td>0.68</td>
</tr>
<tr>
<td>S10</td>
<td>62.3</td>
<td>97.4</td>
<td>31.17</td>
<td>1938</td>
<td>1208</td>
<td>604</td>
<td>0.64</td>
</tr>
<tr>
<td>S11</td>
<td>76.3</td>
<td>96.5</td>
<td>53.51</td>
<td>1538</td>
<td>1175</td>
<td>823</td>
<td>0.66</td>
</tr>
<tr>
<td>S12</td>
<td>75.5</td>
<td>97.1</td>
<td>73.24</td>
<td>1196</td>
<td>904</td>
<td>876</td>
<td>0.61</td>
</tr>
<tr>
<td>S13</td>
<td>84.2</td>
<td>94.3</td>
<td>138.7</td>
<td>1431</td>
<td>1205</td>
<td>1986</td>
<td>0.52</td>
</tr>
<tr>
<td>S14</td>
<td>74.4</td>
<td>94.0</td>
<td>97.65</td>
<td>1614</td>
<td>1201</td>
<td>1576</td>
<td>0.55</td>
</tr>
<tr>
<td>S17</td>
<td>77.9</td>
<td>96.8</td>
<td>70.95</td>
<td>1191</td>
<td>928</td>
<td>845</td>
<td>0.63</td>
</tr>
<tr>
<td>S18</td>
<td>73.9</td>
<td>95.2</td>
<td>78.33</td>
<td>1680</td>
<td>1243</td>
<td>1316</td>
<td>0.59</td>
</tr>
<tr>
<td>S19</td>
<td>78.9</td>
<td>96.6</td>
<td>79.77</td>
<td>1058</td>
<td>835</td>
<td>844</td>
<td>0.61</td>
</tr>
<tr>
<td>All</td>
<td>71.2</td>
<td>96.73</td>
<td>64.27</td>
<td>16945</td>
<td>12080</td>
<td>10890</td>
<td>0.61</td>
</tr>
</tbody>
</table>

4.3 K complex detection using clustering method

Unlike the results for automatic sleep spindle detection, the results for K-complex detection were only evaluated on one database since it was the only public domain database available.

4.3.1 Dreams K-complex database

Results using the clustering method described in sub-section 3.3.2 were obtained for different window sizes and plotted in the figures shown. An ROC curve plot was not possible as the Sensitivity vs FPP did not show a clear relationship. A window length of 0.9 seconds showed the best results in terms of F₁-score, this can be seen in Fig. 4-12 showing F₁-scores for different window sizes and Table 4-15. Window sizes larger and smaller than 0.9 seconds showed a decreasing F₁-score. The results obtained using the clustering method for this database were poorer than results from existing studies discussed in section 7.1.2 and sub-section 7.1.2.1.
Figure 4-12: $F_1$-score results obtained using different window sizes on the Dreams K-complex database.

![Graph showing $F_1$-score vs Moving Window Length]

Table 4-15: Sensitivity, FPP and $F_1$-score results for different window sizes

<table>
<thead>
<tr>
<th>Moving window length (seconds)</th>
<th>Sensitivity %</th>
<th>FPP %</th>
<th>$F_1$-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>44</td>
<td>67</td>
<td>0.4158</td>
</tr>
<tr>
<td>0.6</td>
<td>52</td>
<td>100</td>
<td>0.411</td>
</tr>
<tr>
<td>0.7</td>
<td>63</td>
<td>109</td>
<td>0.466</td>
</tr>
<tr>
<td>0.8</td>
<td>59</td>
<td>106</td>
<td>0.4472</td>
</tr>
<tr>
<td>0.9</td>
<td>75</td>
<td>104</td>
<td>0.538</td>
</tr>
<tr>
<td>1</td>
<td>70</td>
<td>102</td>
<td>0.516</td>
</tr>
<tr>
<td>1.1</td>
<td>69.16</td>
<td>101</td>
<td>0.511</td>
</tr>
<tr>
<td>1.2</td>
<td>70</td>
<td>100</td>
<td>0.517</td>
</tr>
<tr>
<td>1.3</td>
<td>66</td>
<td>105</td>
<td>0.491</td>
</tr>
<tr>
<td>1.4</td>
<td>66</td>
<td>103</td>
<td>0.495</td>
</tr>
<tr>
<td>1.5</td>
<td>65</td>
<td>102</td>
<td>0.487</td>
</tr>
<tr>
<td>1.6</td>
<td>66</td>
<td>100</td>
<td>0.497</td>
</tr>
<tr>
<td>1.7</td>
<td>65</td>
<td>103</td>
<td>0.488</td>
</tr>
<tr>
<td>1.8</td>
<td>65</td>
<td>106</td>
<td>0.48</td>
</tr>
<tr>
<td>1.9</td>
<td>60</td>
<td>104</td>
<td>0.455</td>
</tr>
<tr>
<td>2</td>
<td>60</td>
<td>102</td>
<td>0.459</td>
</tr>
<tr>
<td>2.1</td>
<td>62</td>
<td>103</td>
<td>0.467</td>
</tr>
<tr>
<td>2.2</td>
<td>58</td>
<td>101</td>
<td>0.447</td>
</tr>
<tr>
<td>2.3</td>
<td>56</td>
<td>107</td>
<td>0.429</td>
</tr>
<tr>
<td>2.4</td>
<td>59</td>
<td>109</td>
<td>0.441</td>
</tr>
<tr>
<td>2.5</td>
<td>51</td>
<td>107</td>
<td>0.4</td>
</tr>
</tbody>
</table>
4.4 K complex detection using Pattern matched wavelet.
This section details the results obtained using the pattern matched wavelet technique described in section 3.3.3.1. A seed K-complex was used to construct a wavelet which was then used to detect all K-complexes via the continuous Wavelet transform.

4.4.1 Dreams K-complex database
Table 4-16 shows the results for the Dreams K-complex database using a matched wavelet #1 (shown in Fig. 3-4). An overall sensitivity of 84% was observed at an FPP of 50.87%. Various thresholds ranging from 100 to 800 were used to assess the results and plotted in Fig. 4-13. A Threshold of 400 was found to produce the best results. Fig. 4-6 shows the changes in sensitivity vs precision curves for different thresholds. Results using the Pattern matched wavelet method showed better performance in terms of sensitivity and FPP and F1-score when compared with the clustering method in the previous section. Detailed discussion of these results is presented in sub-section 7.1.2.

Table 4-16 Results obtained using Pattern matched wavelet #1. Overall Sensitivity of 84% at a Precision of 62% was observed.

<table>
<thead>
<tr>
<th>Subject</th>
<th>True Positives</th>
<th>False Positives</th>
<th>Total # of K-complexes scored by visual scorers</th>
<th>Sensitivity %</th>
<th>Precision %</th>
<th>FPP %</th>
<th>F1-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>28</td>
<td>18</td>
<td>44</td>
<td>63</td>
<td>60</td>
<td>40.90</td>
<td>0.61</td>
</tr>
<tr>
<td>2</td>
<td>45</td>
<td>26</td>
<td>45</td>
<td>100</td>
<td>63</td>
<td>57.77</td>
<td>0.77</td>
</tr>
<tr>
<td>3</td>
<td>8</td>
<td>15</td>
<td>12</td>
<td>66</td>
<td>34</td>
<td>125</td>
<td>0.44</td>
</tr>
<tr>
<td>4</td>
<td>67</td>
<td>29</td>
<td>82</td>
<td>81</td>
<td>69</td>
<td>35.36</td>
<td>0.74</td>
</tr>
<tr>
<td>5</td>
<td>45</td>
<td>28</td>
<td>45</td>
<td>100</td>
<td>61</td>
<td>62.22</td>
<td>0.75</td>
</tr>
<tr>
<td>All</td>
<td>193</td>
<td>116</td>
<td>228</td>
<td>84</td>
<td>62</td>
<td>50.87</td>
<td>0.71</td>
</tr>
</tbody>
</table>
Results obtained using matched wavelet #2 (shown in Fig. 3-5) are given in Table 4-17. An overall sensitivity of 74% was obtained at an FPP of 42.98%. A threshold of 400 was used.

Table 4-17 Results obtained using pattern matched wavelet #2, Overall Sensitivity of 74% at a Precision of 42.9% was observed.

<table>
<thead>
<tr>
<th>Subject</th>
<th>True Positives</th>
<th>False Positives</th>
<th>Total # of K-complex's scored by visual scorers</th>
<th>Sensitivity %</th>
<th>Precision %</th>
<th>FPP %</th>
<th>F-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>25</td>
<td>15</td>
<td>44</td>
<td>56</td>
<td>62</td>
<td>34.09</td>
<td>0.59</td>
</tr>
<tr>
<td>2</td>
<td>34</td>
<td>21</td>
<td>45</td>
<td>75</td>
<td>61</td>
<td>46.67</td>
<td>0.67</td>
</tr>
<tr>
<td>3</td>
<td>12</td>
<td>11</td>
<td>12</td>
<td>100</td>
<td>52</td>
<td>91.67</td>
<td>0.68</td>
</tr>
<tr>
<td>4</td>
<td>58</td>
<td>24</td>
<td>82</td>
<td>70</td>
<td>70</td>
<td>29.27</td>
<td>0.70</td>
</tr>
<tr>
<td>5</td>
<td>41</td>
<td>27</td>
<td>45</td>
<td>91</td>
<td>60</td>
<td>60.00</td>
<td>0.72</td>
</tr>
<tr>
<td>All</td>
<td>170</td>
<td>98</td>
<td>228</td>
<td>74</td>
<td>63</td>
<td>42.98</td>
<td>0.68</td>
</tr>
</tbody>
</table>

Figure 4-13 Sensitivity (TPR) vs Precision (PPV) for different detection thresholds on CWT coefficients. Points representing threshold of 400 are marked in the figure. Two lines show different K-complexes chosen for pattern matching i.e matched wavelet #1, matched wavelet #2.
4.5 Summary Tables of Sleep Transient Detection

The following tables show a summary comparison between the different sleep spindle and K-complex detection methods developed in this dissertation. Since the $F_1$-score measure is a single parameter that can be used to compare algorithms, it has been used in these tables.

Table 4-18, Summary Table showing best results assessed using $F_1$-score obtained using the different sleep spindle detection methods.

<table>
<thead>
<tr>
<th>Method</th>
<th>Dreams Database</th>
<th>MASS database</th>
<th>Charité Database</th>
</tr>
</thead>
<tbody>
<tr>
<td>STFT Method</td>
<td>0.707</td>
<td>0.59</td>
<td>0.44</td>
</tr>
<tr>
<td>IIR Filters Method</td>
<td>0.64</td>
<td>0.63</td>
<td>0.46</td>
</tr>
<tr>
<td>Random Forest Classifier</td>
<td>NA</td>
<td>0.61</td>
<td>NA</td>
</tr>
</tbody>
</table>

Table 4-19, Summary Table showing best results assessed using $F_1$-score obtained using the different K-complex detection methods.

<table>
<thead>
<tr>
<th>Method</th>
<th>$F_1$-score for Dreams Database</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clustering method</td>
<td>0.538</td>
</tr>
<tr>
<td>Pattern matched wavelet method</td>
<td>0.71</td>
</tr>
</tbody>
</table>
5 Auditory Stimulation and Neurofeedback for Enhancing Sleep Spindle Count

Using the SS detection technique (clustering method) developed in Chapter 3, a study was conducted to increase SS count using two different methods. One method was through the use of auditory stimulation during sleep and the other method was using NF prior to sleep. The results from both of these methods along with the results from the combination of the two methods were analysed and compared post study. This chapter details the study design, setup and methodology.

5.1 Study Design
The study was designed to include six participants in three different groups. Participants were recruited through advertisements. The study was approved by the CHEAN (College Human Ethics Advisory Network) ethics committee which is a subcommittee of the HREC (Human Research Ethics Committee) at the Royal Melbourne Institute of Technology (RMIT University).

On initial email/telephone contact in response to the advertisement, potential participants were given a brief description of the experiment. Participants were screened to ensure no prior history of drug or alcohol abuse and neurological, psychiatric or sleep disorders. Participants were not considered habitual nappers based on a sleep habit questionnaire obtained at the initial screening process, indicating one or less naps per week on average. Written informed consent was obtained from all participants prior to their participation. Participants were asked to maintain a regular sleep schedule one week prior to the study and abstain from caffeine, non-experimental naps, and alcohol throughout the course of the study day.

5.1.1 Participants:
6 participants between the ages of 18 and 45 were randomly assigned to either a

1) Neurofeedback prior to napping group;

2) Auditory stimulation during napping group; or

3) Neurofeedback prior to napping + auditory stimulation during napping group.

5.1.2 Methodology:

Day 1:
Baseline measurements were taken on this day. Participants in all groups undertook a nap for a 90 minute period. The EEG of the nap was recorded. Additionally, participants in the NF groups 1 and 3 undertook an NF session after their nap on Day 1 in order to accustom them to the NF sessions that they will undertake on Days 2 - 10.
Days 2 - 10:

Groups 1 and 3 undertook 9 sessions from day 2 to 10 whereas Group 2 (auditory stimulation group) only undertook 4 sessions during those corresponding sessions. The discrepancy in the number of sessions between Groups 1 and 3 and Group 2 is because as suggested in literature, SMR NF can take up to multiple sessions to produce significant results. However, there is no necessity for Group 2 to do similar number of sessions. This allowed us to cut down the number of total sessions required in this pilot study.

Participants in NF Groups 1 and 3 will undertook an NF session lasting approximately 50 minutes followed by the word pair association memory task. In contrast, participants in Group 2 only undertook the word pair association task prior to their nap.

Following the NF and word pair association memory tasks, all groups attempted a 90-minute nap in our lab. Groups receiving auditory stimulation (Groups 2 and 3) during napping had low decibel auditory tones (50dB +/- 5dB) played to them during NREM sleep stages 2 and 3. All groups further undertook a final word pair association task to complete the experiment.

EEG Recording of the Napping Period:

Daytime napping EEG data was collected using a g-MOBILAB (g.tec, Austria) PSG data acquisition device with data viewer and signal analysis software. This is medical grade recording equipment that meets major European standards, i.e., the CE Mark [108].

Memory Consolidation Tasks:

All participants on days 1-10 undertook two memory consolidation tasks, prior to their nap and after their nap. This is based on similar experiments conducted in [18] and [89]. Further details of the task are described in sub-section 5.2.1. Memory consolidation will be measured based on the difference in the number of words recalled in the tests before and after the nap.

Data Analysis:

The memory consolidation was measured based on the differences in results obtained before and after the nap. Sleep spindle density for different groups were calculated and compared. Further correlations between spindle density and memory consolidation were calculated. ANOVA (analysis of variance) was used to assess overall differences amongst the three groups, p-values less than 0.05 were considered to be statistically significant [123].

5.2 Materials Design and Apparatus

5.2.1 Declarative Memory Task: Paired Associate Learning

To test declarative memory changes related to SS count and density, a word pair association task was used. Other researchers in the past have used similar word pair association tasks [89][18][109][110]. The word pair association task was designed to measure declarative memory changes in experiments. Participants were presented 40 pairs of nouns on a monitor, each for 4 seconds. The words were
semantically related (‘solution – problem’ is an example pair). After the presentation of the list, participants undertook a cued recall test in which they were asked to recall the second word when showing the first word of each pair. Immediately following their response, the correct answer is shown for 4 s. The cued recall test was also performed after the nap, this time without the correct answer being shown following their response. Memory consolidation was measured based on the difference in the number of words recalled in the tests before and after the nap.

5.2.2 Auditory stimulation protocol

Auditory Tone Delivery:

Tones were delivered to the participant 1 second after a real-time detection system detected a spindle. The stimulus was presented using in-ear headphones. The time duration of the tone was set to be maintained at 50 msec duration and the volume at 50dB +/- 5dB. The tone frequency was randomised using custom software deployed on an Arduino Uno board which received triggers from Simulink (MATLAB 2015). The output was then fed into an analogue filter with a -3db/octave slope (pink noise with equal amount of energy across all frequencies). The analogue filter circuit is shown in Fig. 5-1. The filter is based on multiple lag compensators across the 20-15000 Hz range to maintain a 3db/octave slope [111]. The output volume for different frequencies was checked using a Digitech QM1589 digital sound meter. The output volume across the 20-15000 Hz range was calibrated to 50dB. All frequencies showed a constant volume with a +/- 5db maximum variance across multiple tests.

![Figure 5-1 Filter to generate pink noise powers](image)

The transfer function of the circuit is:

\[
H(s) = \frac{2.406 \times 10^{56} s^3 + 9.501 \times 10^{62} s^2 + 3.473 \times 10^{66} s + 1.053 \times 10^{69}}{5.398 \times 10^{54} s^4 + 1.001 \times 10^{60} s^3 + 1.192 \times 10^{67} s^2 + 1.348 \times 10^{69} s + 1.053 \times 10^{69}}
\] (58)

The frequency response of the circuit is shown in Fig. 5-2.
Auditory stimuli were presented binaurally using in ear headphones. Auditory stimulus was timed to be delivered around 1-2 seconds after real-time detection of a sleep spindle. No auditory stimulation was provided during the first baseline recording. Each subject’s first baseline recording was used to define the spindle and non-spindle cluster parameters. These cluster parameters were used to detect spindles in real-time for the rest (following) of the sleep sessions for that subject.

The delay induced by the tone generating circuit was tested and was observed to be 50ms. The Matlab Simulink platform was used to design the real-time spindle detection and auditory tone trigger signal. Order 8 IIR filters (digital Butterworth filters implemented in Simulink) were used to filter out necessary bands according to the filters defined in sub-section 3.2.3.4. A USB serial port was used to transmit auditory triggers from Simulink to the tone generating circuit. The transfer function and the frequency response of the IIR filters are as follows.
$B_4(10.5-15\text{Hz})$ filter:

$$H(s) = \frac{1.75e-05 \, s^8 - 7e-05 \, s^6 + 0.000105 \, s^4 - 7e-05 \, s^2 + 1.75e-05}{s^8 - 7.263 \, s^7 + 23.44 \, s^6 - 43.89 \, s^5 + 52.11 \, s^4 - 40.18 \, s^3 + 19.65 \, s^2 - 5.573 \, s + 0.7025} \tag{59}$$

**Figure 5.3** Frequency response of the 10.5-15 Hz filter

$B_3(4-10 \text{ Hz})$ filter:

$$H(s) = \frac{2.442e-05 \, s^8 - 9.768e-05 \, s^6 + 0.0001465 \, s^4 - 9.768e-05 \, s^2 + 2.442e-05}{s^8 - 7.524 \, s^7 + 24.86 \, s^6 - 47.11 \, s^5 + 56.01 \, s^4 - 42.78 \, s^3 + 20.5 \, s^2 - 5.635 \, s + 0.683} \tag{60}$$
Figure 5-4 Frequency response of the 4-10 Hz filter

\[ B_4(20-40 \text{ Hz}) \text{ filter:} \]

\[
H(s) = \frac{0.002057 s^4 - 0.008228 s^6 + 0.01234 s^4 - 0.008228 s^2 + 0.002057}{s^8 - 5.135 s^7 + 12.7 s^6 - 19.38 s^5 + 19.88 s^4 - 14 s^3 + 6.624 s^2 - 1.935 s + 0.273}
\]  

(61)

Figure 5-5 Frequency response of the 20-40 Hz filter
5.2.3 EEG recordings

All EEG recordings during NF and sleep were recorded using g.Mobilab (biosignal amplification unit, g.tec medical engineering GmbH, Austria) equipment. EEG was recorded for the C3 and C4 recording sites from the 10-20 system (Fig. 3-2). The ground electrode was placed on the forehead and the Cz, C3, channels were referenced to Mastoids M1 and M2. Ag/Au cup electrodes [6] were used in all recordings. All recordings were sampled at 256 Hz (the only sampling frequency available on g.Mobilab) and stored on a PC. Auditory stimulation time locations were acquired using a stimulus trigger box with g.Mobilab digital input line and stored along with the EEG data.

5.2.4 Neurofeedback Protocol

As described in the earlier chapters, SMR NF involves applying feedback of the 12-15 Hz EEG band activity observed in the central regions (corresponding to the sensorimotor cortex) of the brain. The participants undertaking SMR NF are typically advised to try and increase SMR activity shown to them on a computer monitor. This information is presented in the form of vertical moving bars which represent the SMR activity. The SMR NF protocol described here followed a similar protocol used by Kober et al. in 2015 [24].

Each subject undertook multiple SMR NF sessions over a period of 4 weeks. The SMR NF task was followed by the first cued recall word-pair association task and preceded the 90-minute nap. For SMR NF, the subjects were seated in front of a computer at a comfortable distance of their choosing. EEG data was recorded using Simulink (Matlab 2015b) and g.Mobilab (biosignal amplification unit, g.tec medical engineering GmbH, Austria) equipment. Once the EEG electrodes were attached to the subject, they proceeded with the SMR NF blocks. A whole SMR NF session consisted of ten 170 second feedback blocks with 20 second breaks in between blocks (Fig. 5-3). EEG sampling rate was
set to 256 Hz for all subjects and all sessions. The C channel was used to calculate the SMR band (12-15) power and provide feedback.

![Figure 5-7 SMR feedback intra-session blocks, a 20 second break between the 170 second periods is presented using a white cross on a black background.](image)

The SMR feedback was given via a moving green vertical bar (Fig. 5-7, Fig. 5-8). Two other bars were also shown to the participant, including a red bar corresponding to eye movement artifacts (4-7 Hz) and a blue bar corresponding to EMG artifacts (20-40 Hz). Butterworth IIR filters (8th order) were used to calculate real-time band powers for each bar. Points were awarded to the participant when a green bar crossed (moved above) a threshold for longer than 1 second while the red and blue bars were below a threshold for the same duration. These thresholds were calculated based on the average band activity in the preceding 170 second block. The SMR band (Green bar) threshold was set to 1.1 times the average band (12-15 Hz) power in a moving 1 second window in the preceding block. Threshold for the red bar (eye movement artifacts band) was set to 1.2 times and the threshold for the blue bar (EMG artifacts band) was set to 1.2 times as well. The computer screen was updated at a rate of 10 Hz to keep smooth movement of the bars.

![Figure 5-8 SMR feedback screen, the green bar represents SMR activity, while the red and blue bars represent artifact activity which the subject is instructed to try and keep low.](image)
5.2.5 Auditory stimulation + neurofeedback protocol
In the experiment using auditory stimulation plus NF protocol, subjects undertook NF similar to the NF protocol and additionally they were also given auditory stimulation during sleep similar to the auditory stimulation group.

5.2.6 Electroencephalography analyses
Automatic spindle detection was conducted for all subjects and all sessions using the clustering technique described in sub-section 3.2.2 with features described in sub-section 3.2.2.2.
6 Auditory Stimulation and Neurofeedback

Results for each subject were analysed and tabulated. Total number of spindles and spindle density for each subject were calculated using the IIR filters method described in section 3.2. Fig. 6-1 shows the changes in spindle density compared with the baseline for all groups. Baseline spindle density was obtained from the first baseline recording session which every subject undertook as their first session. One-way ANOVA (Appendix B) for the groups showed a significant difference between groups assuming a 0.05 significance level ($F = 6.2, p < 0.0054$). Post hoc analysis did not show any difference between the Auditory Stimulation group and the auditory stimulation + NF group. Only the NF group showed lower spindle density compared with the baseline. Further discussion is provided in section 7.2.

![Change in spindle density from baseline for all groups.](image)

Figure 6-1 Change in spindle density from baseline for all groups. The auditory stimulation groups (1 and 2) show a positive change from baseline, whereas the NF only group (3) shows a negative change of baseline. This is contradictory to some findings by other researchers who observed an increase; this is discussed further in Chapter 7.

Fig. 6-2 shows the changes in memory consolidation improvement (MCI) for all groups. One-way ANOVA for the groups showed a significant difference between groups assuming a 0.05 significance level ($F = 4.71, p < 0.018$). Post hoc analysis showed no significant difference between the auditory stimulation + NF group and the NF group. The auditory stimulation only group showed significant changes compared with other groups. It is interesting to note that the auditory stimulation group was the only group which showed higher MCI compared with the baseline, implying better memory consolidation compared with the baseline.
Figure 6-2 Changes in the memory consolidation parameter compared to baseline for all groups. Only Group 1 showed a positive change, whereas both the NF groups (2 and 3) show a negative change.

Tables 6-1 to 6-6 contain results for individual subjects. The results include SS count, SS density, change in SS density from baseline and the MCI parameter for every session undertaken by the subject.

MCI shown in the tables stands for memory consolidation improvement, and is defined as

\[ MCI = N_{\text{pos}} - N_{\text{pre}} \]  

(62)

where \( N_{\text{pos}} \) represents the number of correct word pair solutions in post nap test and \( N_{\text{pre}} \) the number of correct word pair solutions in the pre nap test.

Table 6-1 Memory consolidation and SS improvement analysis – Subject 1- auditory stimulation only protocol

<table>
<thead>
<tr>
<th>Session</th>
<th>1 (baseline)</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total spindles</td>
<td>175</td>
<td>188</td>
<td>246</td>
<td>189</td>
<td>145</td>
</tr>
<tr>
<td>Spindles density</td>
<td>1.9</td>
<td>1.85</td>
<td>2.63</td>
<td>2.34</td>
<td>1.67</td>
</tr>
<tr>
<td>MCI</td>
<td>5</td>
<td>16</td>
<td>8</td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td>Change in spindle density from baseline</td>
<td>NA</td>
<td>-0.05</td>
<td>0.73</td>
<td>0.44</td>
<td>-0.23</td>
</tr>
</tbody>
</table>
### Table 6-2 Memory consolidation and SS improvement analysis – Subject 2 - auditory stimulation only protocol

<table>
<thead>
<tr>
<th>Session</th>
<th>1 (baseline)</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total spindles</td>
<td>323</td>
<td>348</td>
<td>296</td>
<td>312</td>
<td>387</td>
</tr>
<tr>
<td>Spindles density</td>
<td>3.45</td>
<td>3.89</td>
<td>3.47</td>
<td>3.42</td>
<td>3.75</td>
</tr>
<tr>
<td>MCI</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>10</td>
<td>9</td>
</tr>
<tr>
<td>Change in spindle density from baseline</td>
<td>NA</td>
<td>0.44</td>
<td>0.02</td>
<td>-0.03</td>
<td>0.3</td>
</tr>
</tbody>
</table>

Sessions 6 to 10 with this subject involved the subject answering all 40 word-pairs correctly prior to nap and post nap. These were excluded from memory consolidation analysis as improvement consolidation cannot be assessed from those sessions.

### Table 6-3 Memory consolidation and SS improvement analysis – Subject 3 - auditory stimulation + NF protocol

<table>
<thead>
<tr>
<th>Session</th>
<th>1 (baseline)</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total spindles</td>
<td>340</td>
<td>373</td>
<td>328</td>
<td>369</td>
<td>375</td>
<td>408</td>
<td>343</td>
<td>422</td>
<td>450</td>
<td>367</td>
</tr>
<tr>
<td>Spindles density</td>
<td>3.78</td>
<td>4.2</td>
<td>3.7</td>
<td>4.06</td>
<td>3.89</td>
<td>4.2</td>
<td>4.1</td>
<td>4.96</td>
<td>5.48</td>
<td>3.98</td>
</tr>
<tr>
<td>MCI</td>
<td>16</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Average SMR power (micro uV)</td>
<td>NA</td>
<td>0.124</td>
<td>0.126</td>
<td>0.126</td>
<td>0.126</td>
<td>0.126</td>
<td>0.126</td>
<td>0.126</td>
<td>0.126</td>
<td>0.113</td>
</tr>
<tr>
<td>Change in spindle density from baseline</td>
<td>NA</td>
<td>0.42</td>
<td>-0.5</td>
<td>0.36</td>
<td>-0.17</td>
<td>0.31</td>
<td>-0.1</td>
<td>0.86</td>
<td>0.52</td>
<td>-1.5</td>
</tr>
</tbody>
</table>

### Table 6-4 Memory consolidation and SS improvement analysis – Subject 4 - auditory stimulation + NF protocol (Subject could only do 7 sessions)

<table>
<thead>
<tr>
<th>Session</th>
<th>1 (baseline)</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total spindles</td>
<td>216</td>
<td>257</td>
<td>253</td>
<td>273</td>
<td>338</td>
<td>361</td>
<td>232</td>
</tr>
<tr>
<td>Spindles density</td>
<td>2.56</td>
<td>2.86</td>
<td>2.98</td>
<td>3.01</td>
<td>3.98</td>
<td>3.41</td>
<td>2.55</td>
</tr>
<tr>
<td>MCI</td>
<td>12</td>
<td>6</td>
<td>2</td>
<td>12</td>
<td>10</td>
<td>12</td>
<td>5</td>
</tr>
<tr>
<td>Average SMR power</td>
<td>NA</td>
<td>0.124</td>
<td>0.126</td>
<td>0.126</td>
<td>0.124</td>
<td>0.125</td>
<td>0.125</td>
</tr>
<tr>
<td>Change in spindle density from baseline</td>
<td>NA</td>
<td>0.3</td>
<td>0.42</td>
<td>0.45</td>
<td>1.42</td>
<td>0.85</td>
<td>-0.01</td>
</tr>
</tbody>
</table>
Table 6-5 Memory consolidation and SS improvement analysis – Subject 5 - NF only protocol

<table>
<thead>
<tr>
<th>Session</th>
<th>1 (baseline)</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total spindles</td>
<td>385</td>
<td>153</td>
<td>196</td>
<td>309</td>
<td>121</td>
<td>206</td>
<td>318</td>
<td>352</td>
<td>386</td>
</tr>
<tr>
<td>Spindles density</td>
<td>4</td>
<td>1.7</td>
<td>2.3</td>
<td>3.3</td>
<td>1.54</td>
<td>2.7</td>
<td>4</td>
<td>3.9</td>
<td>4.15</td>
</tr>
<tr>
<td>MCI</td>
<td>9</td>
<td>6</td>
<td>6</td>
<td>7</td>
<td>3</td>
<td>-1</td>
<td>3</td>
<td>8</td>
<td>2</td>
</tr>
<tr>
<td>Average SMR power</td>
<td>NA</td>
<td>0.11</td>
<td>0.126</td>
<td>0.126</td>
<td>0.123</td>
<td>0.125</td>
<td>0.126</td>
<td>0.126</td>
<td>0.126</td>
</tr>
<tr>
<td>Change in spindle density from baseline</td>
<td>NA</td>
<td>-2.3</td>
<td>-1.7</td>
<td>-0.7</td>
<td>-2.46</td>
<td>-1.3</td>
<td>0</td>
<td>-0.1</td>
<td>0.15</td>
</tr>
</tbody>
</table>

Table 6-6 Memory consolidation and SS improvement analysis – Subject 6 - NF only protocol (Subjects could only do 4 sessions)

<table>
<thead>
<tr>
<th>Session</th>
<th>1 (baseline)</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total spindles</td>
<td>324</td>
<td>422</td>
<td>235</td>
<td>218</td>
</tr>
<tr>
<td>Spindles density</td>
<td>3.51</td>
<td>4.81</td>
<td>3</td>
<td>2.3</td>
</tr>
<tr>
<td>Memory consolidation improvement</td>
<td>4</td>
<td>3</td>
<td>9</td>
<td>3</td>
</tr>
<tr>
<td>Average SMR power</td>
<td>0.121</td>
<td>0.126</td>
<td>0.119</td>
<td></td>
</tr>
<tr>
<td>Change in spindle density from baseline</td>
<td>1.3</td>
<td>-0.51</td>
<td>-1.21</td>
<td></td>
</tr>
</tbody>
</table>

Tables 6-7 show the results of the post nap tiredness questionnaire. One-way ANOVA showed no significant difference between the groups when tiredness ratings were compared with the baseline for each subject ($F = 0.88, p < 0.4266$). Subject 5 reported being very tired after session 8. Following further analysis, it was observed that subject 5 was woken up in the NREM-3 sleep stage (deep sleep). Subjects waking from the NREM-3 sleep stage have been known to report feelings of tiredness [116], therefore, session 8 from subject 5 was excluded while analysing the post nap questionnaire results using one-way ANOVA.
Fig. 6-3 shows the changes in tiredness ratings compared with the baseline for each group. The tiredness rating scores are presented in Table 6-7. No visually discernible differences in mean can be seen in Fig. 6-3 as suggested by the aforementioned ANOVA.

![Figure 6-3](image.png)

Figure 6-3 Box plots showing changes in tiredness ratings were compared with the baseline. No significant difference was found between any of the groups. Data taken from Tables 6-7

Table 6-7 Post nap questionnaire results. Results show no significant differences from baseline from any of the subjects, these are discussed further in Chapter 7.

<table>
<thead>
<tr>
<th></th>
<th>Tiredness rating (1 to 5), 1 being very tired, 5 being very refreshed</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Subject 1</strong></td>
<td>4 4 4 4 4</td>
</tr>
<tr>
<td><strong>Subject 2</strong></td>
<td>4 4 4 4 4</td>
</tr>
<tr>
<td><strong>Subject 3</strong></td>
<td>3 3 4 4 3</td>
</tr>
<tr>
<td><strong>Subject 4</strong></td>
<td>3 3 4 3 3</td>
</tr>
<tr>
<td><strong>Subject 5</strong></td>
<td>3 4 3 4 3</td>
</tr>
<tr>
<td><strong>Subject 6</strong></td>
<td>3 3 3 3 3</td>
</tr>
<tr>
<td><strong>Session</strong></td>
<td>1 2 3 4 5 6 7 8</td>
</tr>
</tbody>
</table>
6.1 Comparison of first three sessions

Tables 6-8 shows a comparison of the baseline measurement session to the first and the second feedback session in terms total spindle count, spindle density. Fig. 6-4 shows a bar plot of spindle counts during the first three sessions for all subjects and Fig. 6-5 shows a bar plot of spindle density for first three sessions. While subjects #1, and #4 showed an increase from the baseline in sessions 2 and 3, subjects #2 and #3 showed an increase from the baseline only in session 2. Subject #5 belonging to the ‘neurofeedback only’ group showed a decrease from the baseline in sessions 2 and 3. Subject #6, also belonging to the NF group, showed an increase in session 2 and a decrease in session 3. Any clear conclusions could not be made from just 3 sessions and would require more subjects for conclusive evidence.

Table 6-8 Sleep spindle count and density (spindles per minute) of first three sessions for all subjects.

<table>
<thead>
<tr>
<th>Subject</th>
<th>Session 1 (baseline)</th>
<th>Session 2</th>
<th>Session 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Spindle Count</td>
<td>Spindle Density</td>
<td>Spindle Count</td>
</tr>
<tr>
<td>1</td>
<td>175</td>
<td>1.9</td>
<td>188</td>
</tr>
<tr>
<td>2</td>
<td>323</td>
<td>3.45</td>
<td>348</td>
</tr>
<tr>
<td>3</td>
<td>340</td>
<td>3.78</td>
<td>373</td>
</tr>
<tr>
<td>4</td>
<td>216</td>
<td>2.56</td>
<td>257</td>
</tr>
<tr>
<td>5</td>
<td>385</td>
<td>4</td>
<td>153</td>
</tr>
<tr>
<td>6</td>
<td>324</td>
<td>3.51</td>
<td>422</td>
</tr>
</tbody>
</table>

Figure 6-4 Total spindle count of the first three sessions for all subjects. Acronym definitions, A – Auditory Stimulation, NF – Neurofeedback.
Figure 6-5 Sleep spindle density (spindles per minute) of first three sessions for all subjects. Acronym definitions, A – Auditory Stimulation, NF – Neurofeedback.
7 Discussions and Conclusions

7.1 Automatic sleep transient detection

7.1.1 Sleep spindle detection

7.1.1.1 Comparison of the STFT and IIR filter methods

The STFT method and IIR filter method were compared for the different databases using \(F_1\)-scores and AUC (area under the curve) of ROC (receiver operating characteristic) curves. Initially, the STFT method was optimised to each database and the ideal parameters found for each database were different (Dreams database – 1.2 second moving window, Mass database – 1 second moving window, Charité database – 1 second moving window). When the second method, i.e., the IIR filter method was evaluated on the databases, a single 1 second window was used to evaluate all databases to avoid overfitting of parameters to each database.

Comparing the AUC curves for the Dreams database, Fig. 4-7 showed better performance of the STFT method. Results were only compared using 5 subjects of the Dreams database, as one subject had to be excluded due to low sampling frequency. Comparing AUC curves for the longer (whole night recordings) and larger set of the MASS database (with more subjects) and the Charité database, the results showed higher AUC using the IIR filters method. A t-test comparing subject by subject \(F_1\)-scores showed no statistical significance between the IIR filter method and the STFT method for the MASS database \((p>0.05)\).

The processing time required by the IIR filter method was lower compared with the STFT method when analysing whole night recordings (88.38 seconds vs 147.48 seconds). Given the performance of the IIR filters method was better (higher AUC) than the STFT method for larger databases, the lower processing time using the IIR filter method would make it more suitable for longer recordings and larger databases as it would significantly reduce processing time.

The Pearson correlation coefficient between the total number of spindles scored by the STFT and IIR filter methods and the total number of visually scored spindles was used to understand how the methods captured inter subject differences. Both methods showed a statistically significant correlation at a 0.05 level suggesting that they were able to capture inter subject differences (STFT method with \(p<0.0162\), IIR filter method with \(p<0.002\)). Under a 0.01 level of statistical significance only the IIR filter method showed significance. The IIR filter method showed a significantly higher correlation (0.71 vs 0.60) than the STFT method in this aspect.

7.1.1.2 Random Forrest classifier method

The Random Forrest classifier produced an overall Sensitivity of 71.29 % at a False Positive proportion of 64.27% for the MASS database. Compared with the STFT and IIR filter methods, the Random Forrest classifier performed poorly in terms of the \(F_1\)-score (0.60 vs 0.63 for the IIR filter). The Pearson correlation coefficient between the total number of spindles scored by the algorithm to
the total number of visually scored spindles showed no statistical significance at the 0.05 level suggesting that this method is unable to capture inter subject differences ($r = 0.5395, p > 0.07$).

7.1.1.3 Comparison of the STFT and IIR filter methods to other methods in literature.

7.1.1.3.1 Dreams Database
A few researchers in the past tested their algorithms on the Dreams Database. Devuyst and colleagues (2011) have tested their algorithm on the Dreams database and achieved a sensitivity of 70.1% at an FPP of 26.44%. The STFT method in this study produced similar results of 70.26% sensitivity at an FPP of 28.25%. The IIR filter method produced a sensitivity of 76.72% at an FPP of 61.54%. It is noted that the IIR filter parameters were not optimised for the Dreams database and further results could only be produced for 5 subjects as the algorithm could not be executed on one subject with low sampling rate. The STFT method was analysed using all 6 subjects and hence could be used to show a truer comparison with other methods. Imitiaz et al., tested a teager energy operator based method on the Dreams database and obtained a sensitivity of 80.3% at a specificity of 97.6% [40]. At a similar sensitivity of 81.5% (using a 1.2 second window), this research obtained a specificity of 97.6%. Nonclerq and colleagues used a bivariate normal model of frequency and amplitude to detect spindles. Their algorithm was assessed on the Dreams database achieving a sensitivity of 71.1% at a specificity of 98.6% [41]. At a similar sensitivity of 72.8% (using the 1.2 second window), this research obtained a specificity of 98.2%. Parekh et al., assessed their own method based on sparse low-rank optimisation on the Dreams database obtaining an $F_1$-score of 0.66 [43]. The optimal $F_1$-score using the STFT method was 0.707. Post the publication of the MASS database in 2014, testing of new algorithms has primarily moved to the MASS database. Further results are discussed in this context in the following section.

7.1.1.3.2 MASS database
The IIR filter method, which performed better than the STFT method based on AUC, was compared with other methods in literature which used the MASS database.

Tsanas and Clifford were among the first researchers to publish results obtained using the MASS database [17]. They had tested six different algorithms in literature along with two of their own algorithms. Results were reported in the form of sensitivity, false discovery rates and $F_1$-scores. Sensitivities in the range of (16.5 – 83.8%) at a false discovery rate range of (49.5-86.5%) ($FDR = FP/(TP+FP)$) were obtained for all methods tested by them. The best method developed by Tsanas and Clifford themselves produced an $F_1$ score of 0.408, sensitivity of 84%, specificity of 90% and $FDR$ of 44.5%. Results obtained in this study using the IIR filter method ($C_1$ channel which had been used for visual scoring) showed a higher $F_1$-score of 0.621 (0.07 standard deviation).

O’Reilly and Neilsen also tested four algorithms from literature on the MASS database [16]. The sigma index detector based on Huppoanean et al., [14] showed the best $F_1$-score of 0.61. This result was obtained when compared with only one of the visual scorers (V5). In this study, an $F_1$-score of
0.628 was obtained at 74.1% sensitivity and a precision of 54.5% when compared with V5 only. Three other methods tested by O’Reilly and Neilsen were based on Molle et al., [117], Devuyst et al., [26], and Ahmed et al., [34].

Durka et al., tested a matching pursuit based algorithm on the MASS database [118]. They obtained a sensitivity of 63.5% at a precision (a.k.a. positive predictive value) of 47% (median F1-score = 0.54). The precision obtained in this study for a sensitivity of 74.1% was 55.4% at a higher median F1-score of 0.61.

Lajnef et al., developed an algorithm using tunable q-factor wavelet transform and morphological component analysis to detect spindles [42]. They tested their algorithm on the MASS database and obtained a sensitivity of 83.18% at an FDR of 39%. An F1-score was not provided, thereby making a direct comparison difficult. IIR filters method in this study showed an FDR of 44.5% at a 74.1% sensitivity indicating that the performance of the method developed by Lajnef et al is likely to have a higher F1-score. The method developed by Lajnef et al., is not a completely automated method since it requires an initial training set of spindles for each subject. Hence the results of their algorithm were highly optimised to the MASS database.

In 2017, Parekh et al., tested four different algorithms on the MASS database [121]. One of the algorithms called the McSleep algorithm was developed by Parekh and colleagues themselves. The best performance was obtained by the McSleep algorithm with an F1-score of 0.62 (0.06 standard deviation). As stated previously, the F1-score obtained by the algorithm (C3-channel) reported in this dissertation was 0.621 (0.07 standard deviation). The other methods evaluated by Parekh and colleagues were the methods developed by Wendt et al., with an F1-score of 0.54 [119], Martin et al., with an F1-score of 0.32 [120], and DETOKS method developed by Parekh et al., in 2015 with an F1-score of 0.60 [121]. The parameters of the McSleep algorithm by Parekh et al., was optimised to the MASS database.

7.1.1.3.3 Slow and fast spindles (MASS)
The larger MASS database allowed examination spindle specific features such as the slow spindle-fast spindle dichotomy. Montages using the frontal derivations (Table 4-8) showed a higher number of slow spindles compared with montages of central and parietal derivations. This finding is in agreement with previous studies which show higher number of slow spindles in the anterior derivations and higher number of fast spindles in the posterior derivations [122][113]. Although there was a slow-fast dichotomy when comparing numbers, a strict visually discernible slow-fast dichotomy was not seen in the frequency distribution of visually scored spindles (Fig. 4-10). Since there were significant differences when comparing frequency distribution of automatically detected spindles using montage 3 (frontal derivations) and the other montages, it appears that the majority of visually scored spindles were fast spindles from the centro-parietal region. The lack of a visually discernible dichotomy of frequencies in the visually scored spindles requires further research by future researchers to establish - if a strict dichotomy is not to be observed or alternatively the visual scoring has missed a number of slow spindles from the frontal derivations.
7.1.1.3.4 Comparison of results of different Montages (MASS database).

Following O’Reilly and colleagues study which previously compared results using different montages, the study reported by this dissertation also examined how results change with different selections of montages for automatic detection of spindles [36]. The databases used for analysis by O’Reilly and colleagues and this study are different. Spindles in the database used by O’Reilly and colleagues were visually scored for the Fz, Cz and Pz locations whereas in this study, the MASS database was visually scored using the C3 channel. Although the C3 channel showed superiority in terms of higher sensitivity, F1-score and lower FPP, one-way ANOVA comparing the C3 with other montages did not show any significant differences. However, a paired t-test to make the same comparison between C3 and other montages showed significant differences in all cases. Further analysis with a few other large databases such as the MASS database may be required before a consensus on the use of montages instead of single channels may be reached.

7.1.1.3.5 Correlation Coefficient

In this study, the Pearson’s correlation was used to study effectiveness of automatic spindle detection in capturing inter subject differences. To the best of the author’s knowledge, the study reported in this dissertation was the first to implement this. Sleep spindles detection research might benefit if future researchers publish results of this statistic when evaluating their algorithms. A high correlation coefficient is useful in studying difference between healthy and abnormal subjects regardless of other performance statistics such as sensitivity, FPP, F1-score etc.

The IIR filter method showed the highest correlation, which was followed by the STFT method and then followed by the random forest classifier. At a significance level of 0.01, only the IIR filter method was statistically significant. At a significance level of 0.05, both clustering methods, i.e., the IIR filter method and the STFT method showed statistical significance.

7.1.1.3.6 NREM3 – spindles (MASS and Charité Databases)

Sleep spindles in the NREM-3 sleep stage were not visually scored for the MASS database and the Charité database. When NREM-3 spindles were excluded from analysis, there was significant change in FPP. A lower FPP was observed (Table 4-9) when NREM-3 spindles were excluded. This lower FPP is more indicative of the true performance of the detection algorithms. While the exclusion of NREM-3 spindles decreased from 59.6% to 40.6% in the MASS database, the results of the Charité database showed a further reduction from 119.74% to 61.86%. A possible cause of this higher reduction is due to the single scorer used to identify SS in the Charité database compared to the two scorers used in the MASS database. The number of SS scored by two scorers is higher than SS scored by a single scorer which will result in lower FPP values.
7.1.1.3.7 Charité Database-
Compared with results from the MASS database and other results from literature discussed in the previous sections, the results obtained on the Charité database showed poorer performance. Although the sensitivity of the Charité database was similar to that of the MASS database, it showed a high number of false positives resulting in a high FPP. As stated in the previous section, when NREM-3 spindles were excluded for analysis, there was a decrease in FPP (119.74% vs 61.84%). Further visual scoring of spindles in the Charité database was only undertaken by one scorer as opposed to two scorers for the MASS database. This potentially contributed to the high FPP when compared with the MASS database post exclusion of NREM-3 spindles.

Overall clustering techniques for sleep spindle detection have produced on par or better results compared to existing literature while simultaneously capturing inter-subject differences.

7.1.2 K-complex detection

Only one Database, i.e., the Dreams K-complex database was available to test the automatic K-complex detection algorithms [47]. In this study, two features were developed to cluster for K-complex detection, the Slope feature and the Low Delta Index (Section 3.3.2). Results were obtained for different window sizes and compared using the F1-score statistic. Best results of sensitivity – 75%, FPP – 104% and F1-score – 0.538 were obtained using a 0.9 second window. The F1-score vs window length plot (Fig 4-12) showed a curve peaking at 0.9 seconds. K-complexes are limited to a 0.5 – 2.0 second length, a 0.9 second window would capture the slope of all K-complexes and is the likely cause of the curve showing a peak at 0.9. Longer windows may be contaminated with multiple slope values.

The Pearson’s correlation coefficient was used to study the sensitivity of the clustering algorithm to inter subject differences. A high Pearson’s coefficient of $r = 0.93 \ (p = 0.0217)$ was observed. The coefficient was also statistically significant at a 0.05 significance level. This is indicative of inter-subject differences being significantly understood using the clustering algorithm, however unlike the larger spindles databases, this result for the K-complex database is limited due to the low number of subjects available.

The pattern matched wavelet method was used to test an alternative k-complex detection algorithm to be compared with the clustering method. Results were obtained for various thresholds and an ideal threshold value of 400 was found (Fig. 4-13). The pattern matched wavelet showed a higher sensitivity of 84% at a lower FPP of 50.87% indicating better performance than the clustering algorithm. The pattern matched wavelet algorithm showed a slightly lower correlation coefficient ($r = 0.9230$) while still being statistically significant ($p < 0.0254$). Results were also obtained using an
alternate matched wavelet #2 matched to a different wavelet (Fig. 3-5). Results using wavelet #2 showed lower sensitivity and lower FPP. However, the F1-scores for both matched wavelet’s were similar (0.68 vs 0.71). Matched wavelet #2 also showed a lower correlation compared to matched wavelet #1 and the clustering method.

7.1.2.1 Comparison of methods to existing K-complex detection literature
As discussed in Chapter 2, unlike sleep spindle detection algorithms, the numbers of K-complex detection algorithms in literature are few. The results from the algorithms reported in this dissertation have been compared to two other studies which have been evaluated on the Dreams K-complex database.

Devuyst et al., evaluated an algorithm using likelihood thresholds on the Dreams Database and showed a sensitivity of 61.7% with respect to visual scorer 1 and a sensitivity of 60.94% with respect to visual scorer 2 [47]. A respective FPP of 19.62% and 181.25% were also shown. Using a union of scoring (results derived from confusion matrix provided by Devuyst and colleagues), a sensitivity of 59.2% at a precision of 76.7% and an F1-score of 0.66 was produced. The F1-score using the clustering algorithm reported in this dissertation was 0.538 and using pattern matched wavelet method was 0.71. When the F1-score statistic is used to compare the different algorithms, the clustering algorithm in this study showed lower performance when compared with the likelihood thresholds methods and the pattern matched wavelet method showed higher performance.

Krohne et al., developed a wavelet based method – detect-k-complexes, and evaluated it on the Dreams K-complex database [51]. Their algorithm detected K-complexes in NREM-2 sleep stage only. Best results from their algorithm showed a sensitivity of 73% at a precision of 65%. Results obtained by the pattern matched wavelet algorithm showed higher sensitivity (84%) at a similar precision of 63%. Pattern matched wavelet algorithm was not restricted to the NREM-2 sleep stage which resulted in a lower precision. The clustering algorithm showed a precision of 41% at a sensitivity of 75%, also not restricted to the NREM-2 sleep stage.

7.2 Auditory stimulation and neurofeedback protocol
Three groups of 2 subjects each were used to conduct a pilot study to examine the effects of NF prior to sleep and auditory stimulation synchronised to inter-spindle period (which is between two consecutive spindles) during sleep.

7.2.1 Comparison of the three groups
One-way ANOVA of the change in spindle density from baseline for all groups showed a statistically significant difference (Fig. 6-1). Post hoc analysis showed an increase in sleep spindles observed in the ‘Auditory stimulation’ and ‘Auditory stimulation + Neurofeedback’ groups while a negative effect was seen in the NF group. This strongly indicates auditory stimulation increasing the number of the sleep spindles. This is in line with earlier findings of increase in sleep spindles due to auditory
stimulation [18][19][89][90]. Ngo et al., used auditory stimulation which was synchronised to slow oscillations, and Sato et al., manually triggered auditory stimulation during NREM-2 sleep stage [25][91]. During 90 minute naps, Ong et al., used auditory stimulation throughout the NREM-2 and NREM-3 sleep stage, and Antony et al., presented bursts of white noise oscillation to subjects [89][90]. All four studies just mentioned showed an increase in sleep spindles when compared with sham conditions. To the best of the author’s knowledge, the study reported in this dissertation was the first study to attempt synchronisation of auditory stimulation to a post spindle refractory period in order to produce subsequent sleep spindles. Being a pilot study with only 6 subjects, only a basic analysis of total number of spindles was possible.

In the 2014 study conducted by Ngo et al., the sham condition used an auditory stimulation synchronised to the negative peak of the slow oscillations (NREM-3 sleep), thereby establishing that just auditory tones synched to only positive ways showed an increase in slow oscillations and slow oscillation synchronised spindles [18]. False positives in this study may have possibly triggered some spindles in NREM-3 stage, but these are unlikely to be statistically significant given that the tones were not synchronised to the positive peak of slow oscillations. Further studies in the future, which simultaneously test auditory stimulation synchronised to inter-spindle period and auditory stimulation synchronised to slow oscillations in different groups, would shed light on the differences in the two methods.

In 2008, Sato et al., manually applied auditory stimulation during stage 2 sleep and observed a reduction in inter spindle period (minimum period between two consecutive spindles) and higher number of spindles, this is in line with the findings in this dissertation where the process of stimulation has been automated to coincide with the inter spindle period [19].

Similar to a study done by Ngo et al., Ong et al., applied auditory stimulation in synchronisation with positive peaks of slow oscillations in NREM-3 sleep stage [89]. While Ngo et al., tested during overnight sleep, Ong et al., tested during 90-minute day time naps. They showed that the stimulation during day time sleep also showed an increase in sleep spindles and promoted sleep dependant declarative memory (tested using 40 word-pairs similar to the study conducted by this research), where 16 subjects were used in that study. Although the study by this dissertation showed an increase in sleep spindles in the ‘auditory stimulation’ group and the ‘Auditory stimulation + Neurofeedback group’, only the ‘Auditory stimulation’ group showed an increase in memory consolidation compared with baseline. The NF only group showed no improvement or negative improvement compared to baseline on the memory consolidation test. Compared with the ‘Neurofeedback only’ group, the ‘Auditory Stimulation + Neurofeedback group’ showed better performance on memory consolidation parameters compared with the baseline. These facts seemed to indicate a detrimental effect of neurofeedback on memory consolidation and a positive effect of auditory stimulation on memory consolidation. The findings reported in this dissertation are in contradiction to a single blind study conducted in 2008 by Hoedlmoser et al., in which SMR NF was shown to increase declarative memory performance [25]. A similar finding was also presented by two other single blind studies in 2014 by Schabus et al., and in 2015 by Kober et al., [91][24]. However the latest double blind study by Schabus et al., in 2017 did not show any improvements in memory
consolidation or sleep spindles parameters [92]. Double blind studies imply that the researcher conducting analysis of experimental data does not have information on which groups belong to the sham condition or the experimental condition. Although the findings by this research are very preliminary and based on the pilot study data, they are in accordance with latest results from the double blind study by Schabus et al., conducted in 2017. Furthermore, based on the results obtained by this research, NF for day-time naps also seemed to indicate a negative impact on memory consolidation.

Results from the tiredness scale questionnaire (Table 6-7) did not show any significant improvements when compared with the baseline.

Overall the results obtained from auditory stimulation are in line with findings from Sato et al., and the mathematical model of sleep spindles developed by Zygierekicz in 2001 which showed that auditory stimulation during the inter-spindle period would trigger successive spindles [19][60].
7.3 Conclusion

The research presented in this work comprised of multiple studies. The three main studies in this research were the application of clustering techniques to detect sleep spindles, application of clustering techniques to detect K-complexes and the study examining auditory stimulation synchronised to inter-spindle period for enhancement of sleep spindles and memory consolidation.

The first two studies using clustering techniques for detecting sleep spindles and K-complexes were tested on multiple databases and the results indicate that clustering technique using Multivariate Gaussian Mixture Models (MGMM) produced results that were able to capture inter-subject differences; and were on par or better in performance when compared with existing methods in literature. During the process of developing these techniques, novel features that take into account the transient nature of sleep spindles were developed and presented. The contributions to research in this regard are also evidenced in the multiple research publications (List of publications) that resulted from this research. Future research improvements using these methods could be made by incorporating stage-specific features that may help remove false detections in the awake and REM sleep stages. The MGMM clustering approach also appears to be more suited to sleep spindle detection rather than K-complex detection. This is due to the difficulty of developing clustering specific K-complex features since K-complexes are defined primarily by shape, rather than frequency characteristics.

The other main experimental human pilot study conducted showed promising results with respect to auditory stimulation during sleep. Auditory stimulation synchronised to inter-spindle period showed an increase in sleep spindle density and memory consolidation in the strictly auditory stimulation subjects. Since a preliminary case has been established by this work for auditory stimulation synchronised to inter-spindle period, future research can be undertaken in clearly establishing this with the use of a sham-group along with the experimental group. The sham group would be designed to receive auditory stimulation randomly played through a 90-minute nap. Further research can also be conducted in comparing results between groups receiving auditory stimulation synchronised to positive peaks of slow oscillations and groups receiving auditory stimulation synchronised to inter-spindle period.

A summary of the hypotheses tested (section 1.4) in this dissertation is provided here.

*Hypothesis 1 – Automatic detection of sleep spindles using clustering methods without the use of thresholds is an effective detection method to observe inter subject difference.*

As evidenced by statistical tests, automatic detection of sleep spindles using clustering methods was found to be a significantly effective method to observe inter subject differences. The results obtained using the clustering methods were also on par with existing techniques in literature.
**Hypothesis 2** – Automatic detection of K-complexes using clustering methods without the use of thresholds is an effective detection method to observe inter subject differences.

Clustering method to detect K-complexes was shown to be an effective method to observe inter subject differences. This finding is however limited by a low sample size due to the lack availability of public K-complex databases and limited by the poorer overall performance compared to existing methods in literature.

**Hypothesis 3** – Sleep spindle count can be increased and sleep spindles can be entrained with real-time auditory stimulation using automatic sleep spindle detection algorithm developed via a clustering method.

and

**Hypothesis 4** – Increase in sleep spindle count using auditory stimulation is higher than increase in sleep spindle count using SMR neurofeedback protocol.

In this pilot study, sleep spindle count was found to increase due to the entrainment of sleep spindles to auditory stimulation which was synchronised to the inter-spindle period. The synchronisation was performed using an automatic sleep spindle detection algorithm developed using a MGMM clustering technique.

Increase in sleep spindles count using auditory stimulation was found to be higher than increase in sleep spindle count using the SMR neurofeedback protocol. The auditory stimulation only group showed the highest increase followed by a group undertaking combined auditory stimulation and SMR neurofeedback. The SMR neurofeedback group showed the least increase (negative) in sleep spindle count.
Appendix A

This appendix provides details of derivations of the EM algorithm and the convergence of the EM algorithm.

Eq. (30) simplification

The simplification from (30) to (31) is detailed below:

Eq. (30) [100]:

\[
\sum_{y_{1}=1}^{M} \sum_{y_{2}=1}^{M} \cdots \sum_{y_{N}=1}^{M} \sum_{l=1}^{L} \delta_{l,y_{1}} \log(\alpha_{l}p_{l}(x_{i}|\theta_{l})) \prod_{j=1}^{N} p(y_{j}|x_{j},\Theta^{g})
\]

(A1)

\[
\sum_{l=1}^{M} \sum_{i=1}^{N} \log(\alpha_{l}p_{l}(x_{i}|\theta_{l})) \sum_{y_{1}=1}^{M} y_{2}=1 \cdots \sum_{y_{N}=1}^{M} \delta_{l,y_{1}} \prod_{j=1}^{N} p(y_{j}|x_{j},\Theta^{g})
\]

(A2)

\[
\sum_{l=1}^{M} \sum_{i=1}^{N} \log(\alpha_{l}p_{l}(x_{i}|\theta_{l})) \left( \sum_{y_{1}=1}^{M} \cdots \sum_{y_{i-1}=1}^{M} \sum_{y_{i+1}=1}^{M} \cdots \sum_{y_{N}=1}^{M} \prod_{j=1,j\neq i}^{N} p(y_{j}|x_{j},\Theta^{g}) \right) p(l|x_{i},\Theta^{g})
\]

(A3)

\[
\sum_{l=1}^{M} \sum_{i=1}^{N} \log(\alpha_{l}p_{l}(x_{i}|\theta_{l})) \left( \prod_{j=1,j\neq i}^{N} \left( \sum_{y_{1}=1}^{M} p(y_{j}|x_{j},\Theta^{g}) \right) \right) p(l|x_{i},\Theta^{g})
\]

(A4)

\[
\sum_{l=1}^{M} \sum_{i=1}^{N} \log(\alpha_{l}p_{l}(x_{i}|\theta_{l})) \left( \prod_{j=1}^{N} p(y_{j}|x_{j},\Theta^{g}) \right) p(l|x_{i},\Theta^{g})
\]

(A5)

\[
\sum_{l=1}^{M} \sum_{i=1}^{N} \log(\alpha_{l}p_{l}(x_{i}|\theta_{l})) p(l|x_{i},\Theta^{g})
\]

(A6)
Derivations of expectation maximisation update equations (40) and (41)

Substituting (17) in (31):

\[
Q(\Theta, \Theta^\prime) = \sum_{i=1}^{M} \sum_{i=1}^{N} \left( \log(\alpha_i) - \frac{1}{2} \log(|\Sigma_i|) - \frac{1}{2} (x_i - \mu_i)^T \Sigma_i^{-1} (x_i - \mu_i) \right) p(l|x_i, \Theta^\prime)
\]

(A7)

To maximise for $\mu_i$ in (A7), we take the first derivative of (A7) and set it to zero.

\[
\frac{\partial}{\partial \mu_i} \left( \sum_{i=1}^{M} \sum_{i=1}^{N} \left( \log(\alpha_i) - \frac{1}{2} \log(|\Sigma_i|) - \frac{1}{2} (x_i - \mu_i)^T \Sigma_i^{-1} (x_i - \mu_i) \right) p(l|x_i, \Theta^\prime) \right) = 0
\]

(A8)

or

\[
\frac{\partial}{\partial \mu_i} \left( \sum_{i=1}^{M} \sum_{i=1}^{N} \left( - \frac{1}{2} (x_i - \mu_i)^T \Sigma_i^{-1} (x_i - \mu_i) \right) p(l|x_i, \Theta^\prime) \right) = 0
\]

(A9)

Solving (A9), requires the following lemma [101][102]:

**Lemma A1:**

For $A \in \mathbb{R}^{d \times d}$, $z \in \mathbb{R}^{d \times 1}$, \( \frac{\partial (z^T A z)}{\partial z} = (A + A^T)z \)

**Proof:** For $A \in \mathbb{R}^{d \times d}$, $z \in \mathbb{R}^{d \times 1}$

\[
\frac{\partial (z^T A)}{\partial z} = \frac{\partial (A^T z)}{\partial z} = A^T
\]

(A10)

using the product rule of differentiation [101][102]

\[
\frac{\partial (z^T A z)}{\partial z} = (z^T \times \frac{\partial (z^T A)}{\partial z}) + (z^T A \times \frac{\partial z}{\partial z}) = (A + A^T)z
\]

(A11)

Further if $A$ is a symmetric matrix i.e. $A = A^T$.

$(A + A^T)z = 2Az$.

Applying Lemma A1 to (A9):

\[
\sum_{i=1}^{N} \left( \Sigma_i^{-1} (x_i - \mu_i) \right) p(l|x_i, \Theta^\prime) = 0
\]

(A12)

\[
\mu_i = \frac{\sum_{i=1}^{N} x_i p(l|x_i, \Theta^\prime)}{\sum_{i=1}^{N} p(l|x_i, \Theta^\prime)}
\]

(A13)

To maximise for $\Sigma_i^{-1}$, we take the first derivative of (A7) with respect to $\Sigma_i^{-1}$ and set it to zero

\[
\frac{\partial}{\partial \Sigma_i^{-1}} \sum_{i=1}^{M} \sum_{i=1}^{N} \left( \frac{1}{2} \log(|\Sigma_i^{-1}|) - \frac{1}{2} (x_i - \mu_i)^T \Sigma_i^{-1} (x_i - \mu_i) \right) p(l|x_i, \Theta^\prime) = 0
\]

(A14)
Lemma A2: For $A \in \mathbb{R}^{d \times d}$, $z \in \mathbb{R}^{d \times 1}$, $z^T A z = tr(Azz^T)$, where $tr$ is the trace of a matrix.

Proof: Assume $a_{ij}$ is the $i,j$th element of $A$ (element from row $i$ and column $j$) and $z_i$ the $i$th element of $z$.

$$z^T A z = g z$$

where $= z^T A z$, and $g_i$ the $i$th element of $g$. Further $g_i = \sum_{i=1}^{d} z_i a_{ij}$.

$$g z = \sum_{j=1}^{d} g_j z_j = \sum_{j=1}^{d} \sum_{i=1}^{d} z_i a_{ij} z_j$$

$$= \sum_{j=1}^{d} \sum_{i=1}^{d} a_{ij} z_i z_j$$

(A15)

(A16)

(A17)

since $z_i z_j$ represent the $i,j$th element of the matrix $z \times z^T$

$$\sum_{j=1}^{d} \sum_{i=1}^{d} a_{ij} z_i z_j = tr(Azz^T)$$

(A18)

Using Lemma A2 in (A14) results in the following:

$$\frac{\partial}{\partial \Sigma^{-1}_1} \sum_{i=1}^{M} \sum_{i=1}^{N} \left( \frac{1}{2} \log((\Sigma^{-1}_i)) - \frac{1}{2} (r_i - \mu_i)^T \Sigma^{-1}_i (r_i - \mu_i) \right) p(l|x_i, \Theta^R)$$

$$= \frac{\partial}{\partial \Sigma^{-1}_1} \sum_{i=1}^{M} \sum_{i=1}^{N} \left( \frac{1}{2} \log((\Sigma^{-1}_i)) - \frac{1}{2} tr(\Sigma^{-1}_i (r_i - \mu_i)(r_i - \mu_i)^T) \right) p(l|x_i, \Theta^R)$$

$$= \frac{\partial}{\partial \Sigma^{-1}_1} \sum_{i=1}^{M} \sum_{i=1}^{N} \left( \frac{1}{2} \log((\Sigma^{-1}_i)) p(l|x_i, \Theta^R) - \frac{1}{2} tr(\Sigma^{-1}_i (r_i - \mu_i)(r_i - \mu_i)^T) p(l|x_i, \Theta^R) \right)$$

$$= \frac{\partial}{\partial \Sigma^{-1}_1} \sum_{i=1}^{M} \sum_{i=1}^{N} \left[ \frac{1}{2} \log((\Sigma^{-1}_i)) \Sigma^{-1}_i p(l|x_i, \Theta^R) - \frac{1}{2} \Sigma^{-1}_i tr(\Sigma^{-1}_i (r_i - \mu_i)(r_i - \mu_i)^T) p(l|x_i, \Theta^R) \right]$$

(A19)

(A20)

(A21)

Lemma A3:

For $A \in \mathbb{R}^{d \times d}$ and given $A$ is symmetric,

$$\frac{\partial \log|A|}{\partial A} = \begin{cases} \frac{A_{ij}}{|A|} & \text{if } i = j \\ \frac{2A_{ij}}{|A|} & \text{if } i \neq j \end{cases} = 2A^{-1} - \text{diag}(A^{-1})$$

(A22)

where $A_{i,j}$ is the cofactor of $a_{ij}$ ($i,j$th element of $A$) [101][102].
Proof:

The adjoint matrix of $A$ represented $\text{adj}(A)$ is the matrix of all cofactors, $\mathcal{A}_{i,j}$ and $A^{-1} = \frac{\text{adj}(A)}{|A|}$.

Using jacobi’s formula using an arbitrary variable, we have[101][102]:

\[ \frac{\partial |A|}{\partial y} = tr \left( \text{adj}(A) \frac{\partial A}{\partial y} \right) \]  \hspace{1cm} (A23)

\[ \frac{\partial \log |A|}{\partial y} = |A|^{-1} \frac{\partial |A|}{\partial y} = |A|^{-1} \times tr \left( \text{adj}(A) \frac{\partial A}{\partial y} \right) = tr \left( \frac{\text{adj}(A)}{|A|} \frac{\partial A}{\partial y} \right) = tr \left( A^{-1} \frac{\partial A}{\partial y} \right) \]  \hspace{1cm} (A24)

Since $\frac{\partial \log |A|}{\partial A}$ will result in a matrix $\mathcal{R}^{d \times d}$ whose elements are given by $\frac{\partial \log |A|}{\partial a_{ij}}$, we can substitute $a_{ij}$ for $y$ in (A24) yielding,

\[ \frac{\partial \log |A|}{\partial a_{ij}} = tr \left( A^{-1} \frac{\partial A}{\partial a_{ij}} \right) \]  \hspace{1cm} (A25)

Since $A$ is symmetric:

\[ tr \left( A^{-1} \frac{\partial A}{\partial a_{ij}} \right) = \begin{cases} \frac{A_{ii}}{|A|} & \text{if } i = j \\ \frac{2A_{ij}}{|A|} & \text{if } i \neq j \end{cases} \]  \hspace{1cm} (A26)

Substituting (A26) in A(25), we have:

\[ \frac{\partial \log |A|}{\partial A} = \begin{cases} \frac{A_{ii}}{|A|} & \text{if } i = j \\ \frac{2A_{ij}}{|A|} & \text{if } i \neq j \end{cases} = 2A^{-1} - \text{Diag}(A^{-1}) \]  \hspace{1cm} (A27)

Lemma A4:

For $A \in \mathcal{R}^{d \times d}, B \in \mathcal{R}^{d \times d}$ and $A = A^T$(symmetric matrix), it can be shown that

\[ \frac{\partial tr(AB)}{\partial A} = B + B^T - \text{Diag}(B) \]  \hspace{1cm} (A28)

Proof:

\[ tr(AB) = \sum_{i=1}^d \sum_{j=1}^d a_{ij} b_{ij} \]  \hspace{1cm} (A29)
where \( a_{ij}, b_{ij} \) are the elements of the \( A, B \) matrices respectively.

Since \( A \) is a symmetric matrix, \( a_{ij} = a_{ji} \), therefore it follows

\[
\frac{\partial \text{tr}(AB)}{\partial a_{ij}} = \frac{\partial \sum_{i=1}^{d} \sum_{j=1}^{d} a_{ij} b_{ij}}{\partial a_{ij}} = \begin{cases} 
    b_{ij} & \text{if } i = j \\
    2b_{ij} & \text{if } i \neq j
\end{cases}
\]  

(A31)

Since \( \frac{\partial \text{tr}(AB)}{\partial a_{ij}} \) are the elements of the matrix defined by \( \frac{\partial \text{tr}(AB)}{\partial A} \), it follows that:

\[
\frac{\partial \text{tr}(AB)}{\partial A} = B + B^T - \text{Diag}(B)
\]  

(A32)

Using Lemma A3 and Lemma A4 to evaluate derivative in (A21) results in the following:

\[
= \frac{1}{2} \sum_{i=1}^{N} p(l|x_i, \Theta)(2\Sigma_l - \text{Diag}(\Sigma_l)) - \frac{1}{2} \sum_{i=1}^{N} p(l|x_i, \Theta) (2(x_i - \mu_l)(x_i - \mu_l)^T - \text{Diag}((x_i - \mu_l)(x_i - \mu_l)^T))
\]  

(A33)

\[
= \frac{1}{2} \sum_{i=1}^{N} p(l|x_i, \Theta) (2(\Sigma_l - (x_i - \mu_l)(x_i - \mu_l)^T) - \text{Diag}((x_i - \mu_l)(x_i - \mu_l)^T))
\]  

(A34)

Setting the derivative in (A34) to zero implies:

\[
\sum_{i=1}^{N} p(l|x_i, \Theta)((\Sigma_l - (x_i - \mu_l)(x_i - \mu_l)^T)) = 0
\]

\[
\Rightarrow \quad \Sigma_l = \frac{\sum_{i=1}^{N} p(l|x_i, \Theta)(x_i - \mu_l)(x_i - \mu_l)^T}{\sum_{i=1}^{N} p(l|x_i, \Theta)}
\]  

(A35)

This is the update equation stated in (41).
Convergence of EM algorithm

The EM algorithm attempts to iteratively increase the likelihood of a model given some data. The difference between the new likelihood (following and iteration) and the old likelihood (before the iteration) can be written as follows.

\[ L(\theta^{\text{new}}|\mathcal{X}) - L(\theta^{\text{old}}|\mathcal{X}) \]  

(A36)

Since it is easier to deal with log likelihood functions rather than likelihood functions, here we introduce new terminology

\[ K(\theta^{\text{new}}|\mathcal{X}) = \log(L(\theta^{\text{new}}|\mathcal{X})) \quad \text{and} \quad K(\theta^{\text{old}}|\mathcal{X}) = \log(L(\theta^{\text{old}}|\mathcal{X})) \]

where \( \mathcal{X} = \{x_1, x_2, \ldots, x_N\} \) are the set of observed data as stated earlier.

Rewriting \( K(\theta^{\text{new}}|\mathcal{X}) \) in terms of the hidden variable \( y \) yields:

\[ K(\theta^{\text{new}}|\mathcal{X}) - K(\theta^{\text{old}}|\mathcal{X}) = \log \sum_{y \in \mathcal{Y}} P(\mathcal{X}|y, \theta^{\text{new}}) P(y|\theta^{\text{new}}) - \log P(\mathcal{X}|\theta^{\text{old}}) \]  

(A37)

In order to convert (A37) into a more familiar form, we divide and multiply the \( P(\mathcal{X}|y, \theta^{\text{new}})P(y|\theta^{\text{new}}) \) term with \( P(y|\mathcal{X}, \theta^{\text{old}}) \) yielding

\[
= \log \sum_{y \in \mathcal{Y}} P(\mathcal{X}|y, \theta^{\text{new}}) P(y|\theta^{\text{new}}) \times \frac{P(y|\mathcal{X}, \theta^{\text{old}})}{P(y|\mathcal{X}, \theta^{\text{old}})} - \log P(\mathcal{X}|\theta^{\text{old}})
\]

(A38)

\[
= \log \sum_{y \in \mathcal{Y}} P(y|\mathcal{X}, \theta^{\text{old}}) \times \frac{P(\mathcal{X}|y, \theta^{\text{new}})P(y|\theta^{\text{new}})}{P(y|\mathcal{X}, \theta^{\text{old}})} - \log P(\mathcal{X}|\theta^{\text{old}})
\]

(A39)

Since \( \sum_{y \in \mathcal{Y}} P(y|\mathcal{X}, \theta^{\text{old}}) = 1 \), using Jensen’s inequality the following can be proved [103][94][104]:

\[
\log \sum_{y \in \mathcal{Y}} P(y|\mathcal{X}, \theta^{\text{old}}) \times \frac{P(\mathcal{X}|y, \theta^{\text{new}})P(y|\theta^{\text{new}})}{P(y|\mathcal{X}, \theta^{\text{old}})} - \log P(\mathcal{X}|\theta^{\text{old}})
\geq \sum_{y \in \mathcal{Y}} P(y|\mathcal{X}, \theta^{\text{old}}) \times \log \left( \frac{P(\mathcal{X}|y, \theta^{\text{new}})P(y|\theta^{\text{new}})}{P(y|\mathcal{X}, \theta^{\text{old}})} \right) - \log P(\mathcal{X}|\theta^{\text{old}})
\]

(A40)
Rewriting (A40) using (A37)

\[
K(\theta^{\text{new}}|\mathbf{x}) - K(\theta^{\text{old}}|\mathbf{x}) \geq \sum_{y \in Y} P(y|x, \theta^{\text{old}}) \times \log \left( \frac{P(\mathbf{x}|y, \theta^{\text{new}})P(y|\theta^{\text{new}})}{P(y|x, \theta^{\text{old}})} \right) - \log P(x | \theta^{\text{old}})
\]

(A41)

or

\[
K(\theta^{\text{new}}|\mathbf{x}) - K(\theta^{\text{old}}|\mathbf{x}) \geq \Delta(\theta^{\text{new}}|\theta^{\text{old}})
\]

(A42)

where,

\[
\Delta(\theta^{\text{new}}|\theta^{\text{old}}) = \sum_{y \in Y} P(y|x, \theta^{\text{old}}) \times \log \left( \frac{P(\mathbf{x}|y, \theta^{\text{new}})P(y|\theta^{\text{new}})}{P(y|x, \theta^{\text{old}})} \right) - \log P(x | \theta^{\text{old}})
\]

\[
= \sum_{y \in Y} P(y|x, \theta^{\text{old}}) \times \log(P(x|y, \theta^{\text{new}})P(y|\theta^{\text{new}})) - \sum_{y \in Y} P(y|x, \theta^{\text{old}}) - \log P(x | \theta^{\text{old}})
\]

(A43)

From (24) we have : \(Q(\theta, \theta^g) = \sum_{y \in Y} \log(L(\theta|x, y))p(y|x, \theta^g)\). Substituting \(Q(\theta, \theta^g)\) in (A43) we have-

\[
\Delta(\theta^{\text{new}}|\theta^{\text{old}}) = Q(\theta^{\text{new}}, \theta^{\text{old}}) - \sum_{y \in Y} P(y|x, \theta^{\text{old}}) - \log P(x | \theta^{\text{old}})
\]

(A44)

Since the aim is to maximise \(\Delta(\theta^{\text{new}}|\theta^{\text{old}})\) with respect to \(\theta^{\text{new}}\), it implies to maximise the right hand side of (A44).

\[
\theta^{\text{new}} = \arg\max_{\theta^{\text{new}}} Q(\theta^{\text{new}}, \theta^{\text{old}}) - \sum_{y \in Y} P(y|x, \theta^{\text{old}}) - \log P(x | \theta^{\text{old}})
\]

(A45)

Taking the derivative with respect to \(\theta^{\text{new}}\) on the right hand side of (A45) would imply dropping terms constant with respect to \(\theta^{\text{new}}\).

\[
\theta^{\text{new}} = \arg\max_{\theta^{\text{new}}} Q(\theta^{\text{new}}, \theta^{\text{old}})
\]

(A46)

This is the same as the maximisation step of the EM algorithm.

Since the maximisation step implies maximisation of \(\Delta(\theta^{\text{new}}|\theta^{\text{old}})\), and since \(\Delta(\theta^{\text{old}}|\theta^{\text{old}}) = 0\), it implies that \(\Delta(\theta^{\text{new}}|\theta^{\text{old}}) \geq \Delta(\theta^{\text{old}}|\theta^{\text{old}})\) and \(\Delta(\theta^{\text{new}}|\theta^{\text{old}}) \geq 0\). Therefore we have,

\[
K(\theta^{\text{new}}|\mathbf{x}) - K(\theta^{\text{old}}|\mathbf{x}) \geq \Delta(\theta^{\text{new}}|\theta^{\text{old}})
\]

(A47)

\[
K(\theta^{\text{new}}|\mathbf{x}) - K(\theta^{\text{old}}|\mathbf{x}) \geq 0
\]

(A48)

Implies,
\[ L(\Theta^{\text{new}}|X) - L(\Theta^{\text{old}}|X) \geq 0 \]  \hspace{1cm} (A49)

Eq. (A49) implies a strictly increasing likelihood function at each iteration and hence a convergence to a minimum.
Appendix B

Analysis of variance (ANOVA) [123][124]

One way ANOVA is a statistical test used to compare data from different groups and test for the significance in differences. The null-hypothesis for one-way ANOVA assumes that all group means are equal:

\[ H_0: \mu_1 = \mu_2 = \cdots = \mu_k \]  

(B1)

where \( \mu_k \) is the mean of group \( k \). The ANOVA test assumes that data in each group \( k \) is from a normal distribution.

One way ANOVA compares between-group differences and within-group differences as a ratio called the F-statistic or F-ratio. The test also accounts for degrees on freedom available in terms of number of groups and number of data samples. The F-statistic is calculated as follows:

Assume data from \( k^{th} \) group is represented as \( y_{ik} \) and number of data samples in each group is given by \( n_k \)

Individual group means can be represented as follows:

\[ m_k = \frac{1}{n_k} \sum_i y_{ik} \]  

(B2)

Overall mean of all groups is given by:

\[ m = \frac{1}{\sum_k n_k} \sum_k \sum_i y_{ik} \]  

(B3)

The between group sum of squared differences is:

\[ s_b = \sum_k n_k (m_k - m)^2 \]  

(B4)

The between group degrees of freedom is given by \( d_b = (k - 1) \)

The within group sum of squared differences is given by:

\[ s_w = \sum_k \sum_{i=1}^{n_k} (y_{ik} - m_k)^2 \]  

(B5)

The within group degrees of freedom is given by \( d_w = (\sum_k n_k) - k \)

The F-statistic or F-ratio is given by:

\[ F_{ratio} = \frac{s_b/d_b}{s_w/d_w} \]  

(B6)

i.e the F-statistic is a ratio of between-group differences to within-group differences.
The \textit{F-statistic} follows an \textit{F-distribution} defined as follows [124]:

\[ g(F_{\text{ratio}} | d_b, d_w) = \frac{1}{\mathcal{B}(\frac{d_b}{2}, \frac{d_w}{2})} \frac{(d_b)^{\frac{d_b}{2}}}{(d_w)^{\frac{d_w}{2}}} \frac{F_{\text{ratio}}^{\frac{d_b}{2} - 1}}{\left(1 + \frac{F_{\text{ratio}} d_b}{d_w}\right)^{\frac{d_b + d_w}{2}}} \]  

(B7)

Where \( \mathcal{B} \) is the beta function defined as:

\[ \mathcal{B}(a, b) = \int_0^1 z^{a-1} (1 - z)^{b-1} dz \]  

(B8)

The traditional \textit{p-value} is used in estimating significance in ANOVA analysis. The \textit{F-distribution} is right skewed and has support on only positive values, i.e. \( F_{\text{ratio}} \in [0, +\infty) \). The \textit{p-value} for the \textit{F-distribution} is given by \( P(c \geq F_{\text{ratio}} | g) \) where \( c \) is a random variable with distribution \( g(c|d_b, d_w) \) and \( P \) is probability.

For a given significance level such as 0.05, if the \textit{p-value} < 0.05 i.e the probability of group means coming from the same distribution < 5\%, the null hypothesis is rejected.

\textbf{Paired t-test}

This statistical test is used to compare data from two groups with paired data samples i.e. two different methods were applied to the same input data, resulting in two different output results.

The null-hypothesis of the paired t-test assumes that the mean difference between paired observations is zero.

\[ H_0: \mu_1 - \mu_2 = 0 \]  

(B9)

where \( \mu_1, \mu_2 \) are the means of the two groups being compared. The t-test assumes that the data in each group is coming from a normal distribution.

Assume data from the groups are represented as \( y_{1i} \) and \( y_{2i} \) and

The mean difference between the groups is given by:

\[ md = \frac{1}{n} \sum_{i=1}^{n} y_{1i} - y_{2i} \]  

(B10)

The variance of the differences is given by:

\[ vd = \frac{1}{n} \sum_{i=1}^{n} (y_{1i} - y_{2i})^2 \]  

(B11)
The *t-statistic* is defined as follows:

\[
t = \frac{md}{\sqrt{vd/n}}
\]  
(B12)

And the degrees of freedom associated with the *t-statistic* are \(d_f = n - 1\) since there are \(n\) paired data samples.

The *t-statistic* follows a *t-distribution* given by [124]:

\[
g(t|d_f) = \frac{1}{\mathcal{B} \left( \frac{1}{2}, \frac{d_f}{2} \right)} \left( 1 + \frac{t^2}{d_f} \right)^{-(d_f+1)/2}
\]  
(B13)

where \(\mathcal{B}\) is the beta function defined in the previous section.

The *t-distribution* is a symmetric two-tailed distribution; therefore the **p-value** for the t-test is given by \(2 \times \min\{P(c \leq t|g), P(c \geq t|g)\}\) where \(c\) is a random variable with distribution \(g(c|d_f)\) and \(P\) is probability. For a given significance level such as 0.05, if the **p-value** < 0.05 i.e the probability of group means coming from the same distribution < 5%, the null hypothesis is rejected.

**Pearson’s correlation coefficient** [123]

The Pearson’s correlation coefficient or Pearson’s \(r\) is a correlation value between two variables which are paired; i.e. observation in one variable has a corresponding observation in the second variable [applied statistics from bivariate through multivariate techniques].

Assume two observation variables \(X = \{x_1, x_2, \ldots, x_N\}\) and \(Y = \{y_1, y_2, \ldots, y_N\}\). The Pearson’s \(r\) for the two variables is given by:

\[
r_{XY} = \frac{\text{Cov}(X,Y)}{\sigma_X \sigma_Y}
\]  
(B14)

where \(\sigma_X\) and \(\sigma_Y\) are the sample variance of \(X\) and \(Y\), and \(\text{Cov}(X,Y)\) is given by [123]:

\[
\text{Cov}(X,Y) = \frac{\sum_{i=1}^{N} (x_i - \mu_X)(y_i - \mu_Y)}{N - 1}
\]  
(B15)

where \(\mu_X\) and \(\mu_Y\) are the means of \(X\) and \(Y\).

The Pearson’s \(r\) value can be used to test a null-hypothesis which assumes the following:

\[
H_0: r_{XY} = 0
\]  
(B16)
The $r_{XY}$ value follows a t-distribution with N-2 degrees of freedom, therefore t-statistic $r_{XY}$ for and is given by: [123]

$$t = \frac{r\sqrt{N-2}}{\sqrt{1-r^2}}$$  \hspace{1cm} (B17)

As was shown in the earlier section, the t-statistic follows a t-distribution given by [124]:

$$g(t|d_f) = \frac{1}{\mathcal{B}\left(\frac{1}{2}, \frac{d_f}{2}\right)} \left(1 + \frac{t^2}{d_f}\right)^{-\frac{(d_f+1)/2}{d_f}}$$  \hspace{1cm} (B18)

where $d_f = (N - 2)$ are the degrees of freedom and $\mathcal{B}$ is the beta function defined in the previous section.

The t-distribution is a symmetric two-tailed distribution; therefore the p-value for the t-test is given by $2 \times \min\{P(c \leq t|g), P(c \geq t|g)\}$ where $c$ is a random variable with distribution $g(c|d_f)$ and $P$ is probability. For a given significance level such as 0.05, if the p-value < 0.05 i.e. the probability of group means coming from the same distribution < 5%, the null hypothesis is rejected.
Appendix C

Ethics approval letter starting on next page.
5 October 2016

Dr Dean Cvetkovic
School of
Engineering RMIT
University

Dear Dr Cvetkovic

ASEHAPP 63-15 The effects of neurofeedback and auditory stimulation on napping EEG and memory consolidation

Thank you for requesting an amendment to your Human Research Ethics project titled: The effects of neurofeedback and auditory stimulation on napping EEG and memory consolidation, which was originally approved by Science Engineering and Health CHEAN in 2016 for a period of 2 years.

I am pleased to inform you that the CHEAN has approved your amendment as outlined in your request.

The CHEAN notes and thanks you for providing all documentation that incorporates these amendments. This documentation will be appended to your file for future reference and your research may now continue.

The committee would like to remind you that:

All data should be stored on University Network systems. These systems provide high levels of manageable security and data integrity, can provide secure remote access, are backed up on a regular basis and can provide Disaster Recover processes should a large scale incident occur. The use of portable devices such as CDs and memory sticks is valid for archiving; data transport where necessary and for some works in progress; The authoritative copy of all current data should reside on appropriate network systems; and the Principal Investigator is responsible for the retention and storage of the original data pertaining to the project for a minimum period of five years.
Please Note: Annual reports are due on the anniversary of the commencement date for all research projects that have been approved by the CHEAN. Ongoing approval is conditional upon the submission of annual reports failure to provide an annual report may result in Ethics approval being withdrawn.

Final reports are due within six months of the project expiring or as soon as possible after your research project has concluded.

The annual/final reports forms can be found at:  
www.rmit.edu.au/staff/research/human-research-ethics

Yours faithfully,

Associate Professor Barbara
Polus Chair, Science Engineering
& Health College Human Ethics
Advisory Network

Cc    Student Investigator/s:    Chanakya Reddy Patti, School of Engineering
Appendix D
This appendix details the pseudo code/ algorithms described in this dissertation.
Algorithm 1 Expectation Maximisation (EM) for two clusters used in this dissertation using a set convergence threshold of 0.000001

Input: Feature data set or observed feature vectors $\mathcal{X} = \{x_1, x_2, ..., x_N\}$

1. Select two different random observations from $\mathcal{X} = \{x_1, x_2, ..., x_N\}$ as initial cluster centres $\mu_i^\theta$, random covariance matrix of clusters $\Sigma_i^\theta$, and initiate all mixing proportions $\alpha_i^\theta = 0.5$.

2. Calculate log likelihood of model $L(\theta^{old}|\mathcal{X}) = \prod_{i=1}^N p(x_i|\theta)$ given the observed data $\mathcal{X}$

3. Repeat:

   (E-step), Calculate posterior probabilities i.e $p(l|x_i, \Theta)$ of the clusters belonging to each observation in $\mathcal{X}$ using Eq. 26.

   (M-step) Calculate new cluster mixing proportions $\alpha_i^{new}$, cluster centres $\mu_i^{new}$ and cluster covariance matrices $\Sigma_i^{new}$ for each cluster using update Eq.39, Eq.40, Eq.41.

   Calculate log likelihood of the new model $L(\theta^{new}|\mathcal{X}) = \prod_{i=1}^N p(x_i|\theta)$.

Until: Convergence reached, i.e $L(\theta^{new}|\mathcal{X}) - L(\theta^{old}|\mathcal{X}) < 0.000001$.

4. Return final cluster parameters mixing proportions $\alpha_i^{new}$, cluster centres $\mu_i^{new}$ and cluster covariance matrices $\Sigma_i^{new}$.
Algorithm 2 EMG artifacts removal

Input: input signal $S(t)$ (the signal that requires EMG artifacts removal), sampling frequency $F_s$.

Output: Binary signal showing detected spindles as 1’s.

1. $w \leftarrow 5*F_s$
2. Start moving window $W_t$ of length $w$ sample 1 of $S(t)$, i.e. $t = 1$
3. Repeat:
   
   $wrms \leftarrow$ Calculate RMS power in the 20-40 Hz band of $W_t$ using an 8th order IIR filter.
   
   if $wrms > 5uV^2$
   
   Remove signal sample corresponding to $W_t$, i.e. $S(t: t+w) \leftarrow NaN$
   
   end if
   
   $t \leftarrow t+1$

Untill: last sample of $S(t)$ is reached, i.e. start of $W_t = S(t-w)$

4. Return $S(t)$. 

**Algorithm 3** Automatic detection of sleep spindles using STFT and the EM algorithm.

Input: Single channel recording from EEG – \( S(t) \), false spindle check length \( l \), moving window size \( w \).

Output: Binary signal showing detected spindles as 1’s.

1. Remove EMG artifacts in \( S(t) \) using Algorithm 2.
2. Start moving window \( W_i \) at sample \( w + l \) of signal \( S(t) \).
3. Repeat:
   
   Calculate FFT of moving windows starting at \( W_{i-w}, W_i, W_{i+w} \).

   Using FFT’s of \( W_{i-w}, W_i, W_{i+w} \), calculate the Sigma power, Sigma Index, Sigma power 2 and Sigma index 2 features as described in Eqs. 8, 9, 10, 11.

   Move \( W_i \) by one sample, i.e. \( W_i \leftarrow W_{i+1} \)

   Until: last sample of \( S(t) \) is reached, i.e. start of \( W_i = S(t-w) \)

4. Cluster features Sigma power, Sigma Index, Sigma power 2 and Sigma index 2 using EM algorithm described in Algorithm 1.
5. \( O(t) \leftarrow \) Read output of EM algorithm.
6. Starting at \( t = 1 \)
7. Repeat:
   
   Find next spindle in \( O(t) \) by incrementing \( t \).

   Record length of spindle \( s_l \)

   If \( s_l < l \)

   Discard spindle, overwrite spindle location in \( O(t) \) with zero’s

   End If

   Until: last spindle detected in \( O(t) \)

8. Return \( O(t) \) representing all detected spindles.
Algorithm 4 Automatic detection of sleep spindles using IIR filters and the EM algorithm.

Input: Mean of multiple channel (montage) recordings from EEG – $S(t)$, false spindle check length $l$, moving window size $w$.

Output: Binary signal showing detected spindles as 1’s.

1. Remove EMG artifacts in $S(t)$ using Algorithm 2.
2. Filter $S(t)$ using IIR filters into $(10.5 - 15$ Hz) $B_1(t)$, $(4 - 10$ Hz) $B_3(t)$ and $(20 - 40$ Hz) $B_4(t)$
3. Start moving window $W_i$ at sample $w+l$ of filtered signals $B_1(t), B_3(t)$ and $B_4(t)$.
4. Repeat:
   Using $W_{wu}$, $W_i$, $W_{ws}$, of $B_1(t), B_3(t)$ and $B_4(t)$ calculate the Sigma Ratio and Mean Sigma Index as described in Eqs. 15,16.
   Move $W_i$ by one sample, i.e. $W_i ← W_{is+1}$
   Untill: last sample of $S(t)$ is reached, i.e. start of $W_i = S(t-w)$
5. Cluster features Sigma Ratio and Mean Sigma Index using EM algorithm described in Algorithm 1.
6. $O(t) ←$ Read output of EM algorithm.
7. Starting at $t = 1$
8. Repeat:
   Find next spindle in $O(t)$ by incrementing $t$.
   Record length of spindle $sl$
   If $sl < l$
      Discard spindle, overwrite spindle location in $O(t)$ with zero’s
   End If
   Untill: last spindle detected in $O(t)$
9. Return $O(t)$ representing all detected spindles.
Algorithm 5 Calculation of features and training Random Forest Classifier for sleep spindle detection

Input: Training signals $S(t)$ and corresponding visual scorings of sleep spindles $V(t)$, sampling frequency $F_s$

Output: Random Forest Classifier model for sleep spindle detection.

1. Remove EMG artifacts in $S(t)$ using Algorithm 2.
2. Filter $S(t)$ using IIR filters into $(12.5-15 \text{ Hz}) B_1(t)$, $(10.5-16 \text{ Hz}) B_2(t)$, $(4-10 \text{ Hz}) B_3(t)$, $(20-40 \text{ Hz}) B_4(t)$ and $(8-12 \text{ Hz}) B_5(t)$.
3. Start moving window at $t \leftarrow (F_s*1) + 1$
4. Repeat:

   Calculate and store feature Alpha Ratio – $AR(t) \leftarrow \frac{RMS(B_3(t-(F_s+1):t))}{RMS(S(t-(F_s+1):t))}$

   Calculate and store feature Spindle Band Ratio – $SBR(t) \leftarrow \frac{RMS(B_4(t-(F_s+1):t))}{RMS(S(t-(F_s+1):t))}$

   Calculate and store feature Mean Sigma Index described in Eq. 16.

   $t \leftarrow t + 1$

   Until: $(t+(F_s*1)+1) > \text{end of } S(t)$

5. Set vector $x_t \leftarrow [AR(t), SBR(t), M_\sigma (t)]$ as Random Forest Classifier training inputs and training outputs are visual scorings $c_t \leftarrow V(t)$.
6. Send $X = \{x_1, x_2, \ldots, x_n\}$ as input data and $C = \{c_1, c_2, \ldots, c_k\}$ as scorings to Random Forest Training algorithm (Algorithm 5)
7. Read output of Algorithm 5 and store trained model of Random Forest Classifier
Algorithm 6 Training Random Forest Classifier for sleep spindle detection

Input: Training data \( \mathcal{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\} \) belonging to output classes \( \mathcal{C} = \{c_1, c_2, \ldots, c_k\} \), number of features \( F \), and number of trees in forest \( B \)

Output: Random Forest Classifier model for sleep spindle detection.

1. function RandomForest(\( \mathcal{X}, \mathcal{C}, F, B \))
2. \( \mathcal{H} \leftarrow \emptyset \)
3. for \( i \in 1,2,\ldots,B \) do
4. \( S^i \leftarrow \) A bootstrap sample from \( \mathcal{X} \; \mathcal{C} \)
5. \( h^i \leftarrow \) TreeLearn(\( S^i, F \))
6. \( \mathcal{H} \leftarrow \mathcal{H} \cup \{h^i\} \)
7. end for
8. Return \( \mathcal{H} \)
9. end function

10. function TreeLearn(\( S^i, F \))
11. Repeat at each node in \( \mathcal{H} \)
12. \( f \leftarrow \) small random subset of \( F \)
13. Split on best feature in \( f \) using the Gini criterion, Eq. 44.
14. Return learned tree node
15. end function
Algorithm 7 Automatic detection of K-complexes using EM.

Input: Single channel recording from EEG – $S(t)$, moving window size $w$.

Output: Binary signal showing detected K-complexes as 1’s.

1. Remove EMG artifacts in $S(t)$ using Algorithm 2.
2. Start moving window $W_i$ at sample $i$ of signal $S(t)$.
3. Repeat:
   
   Calculate FFT of signal in moving window $W_i$
   
   Calculate the Slope feature $G_K(t)$ for moving $W_i$
   
   Calculate the Low Delta Index $I_{DL}(t) = \int_{0.5}^{2} X_{STFT}(F;t) dF$ feature for moving window $W_i$
   
   Until: last sample of $S(t)$ is reached, i.e. start of $W_i = S(t-w)$

5. $O(t) \leftarrow$ Read output of EM algorithm.
6. Starting at $t = 1$
7. Repeat:

   Find next K-complex in $O(t)$ by incrementing $t$.

   $N \leftarrow$ number of K-complexes in following 20 seconds.

   If $N > 3$
   
   Discard K-complex, overwrite spindle location in $O(t)$ with zero’s
   
   End If

   Until: last K-complex detected in $O(t)$

8. Return $O(t)$ representing all detected K-complexes.
**Algorithm 8** Pattern Matched Wavelet Design and K-complex detection using Pattern Matched Wavelets.

Input: Seed Wavelet SW, Input EEG signal $S(t)$, Threshold for K-complex detection - $Th$, degree of polynomial used for construction of pattern matched wavelet – $d$.

Output: Binary signal showing detected K-complexes as 1’s.

1. $y_k \leftarrow$ Resample SW to 64 samples.
2. Calculate and store polynomial signals $\rho_i$ ($i \in 1,2,...,d$) up to degree $d$, for 64 samples.
3. Calculate $m_i = \int_a^b \rho_i(t)dt$, $g_{ij} = \sum_{k=1}^k \rho_i(t_k)\rho_j(t_k)$, $b_i = \sum_{k=1}^k y_k\rho_i(t_k)$
4. Calculate $\alpha$ using Eq. 53 $(\begin{pmatrix} a \\ m \\ t \end{pmatrix} ) (\begin{pmatrix} \alpha \\ \mu \\ \tau \end{pmatrix} ) = (\begin{pmatrix} b \\ \nu \\ \xi \end{pmatrix} )$
5. Calculate CWT, $X_{CWT}(t,s) \leftarrow \int_{-\infty}^{\infty} S(\tau) \alpha_{t,s}(\tau) d\tau$ for 0.05 seconds $< s < 5$ seconds
6. $t \leftarrow 1$
7. Repeat

   MCWT $\leftarrow \max_{0.5<s<2} X_{CWT}(t,s)$

   if MCWT $> 400$

   Mart true K-complex, $O(t) \leftarrow 1$

   end if

   $t \leftarrow t+1$

   Until: last sample of $S(t)$ is reached

8. $t \leftarrow 1$
9. Repeat:

   Find next K-complex in $O(t)$ by incrementing $t$.

   N $\leftarrow$ number of K-complexes in following 20 seconds.

   If N $> 3$

   Discard K-complex, overwrite spindle location in $O(t)$ with zero’s

   End If

   Until: last K-complex detected in $O(t)$

10. Return $O(t)$ representing all detected K-complexes.
References


