ALCOHOL WITHDRAWAL SYNDROME ASSESSMENT BASED ON TREMOR
TIME-FREQUENCY ANALYSIS

by

Narges Norouzi

A thesis submitted in conformity with the requirements
for the degree of PhD
Graduate Department of Electrical and Computer Engineering
University of Toronto

© Copyright 2017 by Narges Norouzi
Abstract

Alcohol Withdrawal Syndrome Assessment based on Tremor Time-Frequency Analysis

Narges Norouzi
PhD
Graduate Department of Electrical and Computer Engineering
University of Toronto
2017

In this thesis we established signal processing techniques to objectively evaluate the severity of Alcohol Withdrawal (AW) tremors. Many medical protocols were used previously to help the physicians in assessing the severity of alcohol withdrawal, but those techniques were subjective and relied on the experience level of the physicians.

The key objective throughout this thesis is investigating the logarithmic nature of the energy emitted from tremor signals. We are able to use the energy from tremor recordings in the frequency range of $[5, 15]$ Hz to train a logarithmic model to estimate the severity of tremors.

The next step in validating the effectiveness of the logarithmic model is the validation of the methodology in a clinical setting. The model is being validated in the emergency department for a 10-month period. During this period, each of the AW patients have been evaluated by one nurse and have been videotaped while acquiring the signal. The model provides the severity score in realtime after recording the signal. Our model is validated by comparing the score given by the model and the consensus severity rating from a panel of three expert physicians after viewing the videos. We concluded that there is a reliable agreement (kappa 0.92, 95% CI: 0.86, 0.99) between the score given by the model and the rating from our panel.

Further contributions of this thesis include an investigation of the features of AW tremors in classifying factitious vs. real tremors, based on the mean peak frequency and band-limited energy. Additionally we evaluate the differences between AW tremors in
both hands and observe that by averaging the tremor ratings of each hand, a more accurate result can be obtained compared to taking either of the individual hand ratings.

Lastly, to remove the noise from the tremor signal, an Empirical Mode Decomposition (EMD) algorithm was utilized. EMD decomposes the signal into different Intrinsic Mode Functions (IMFs) and IMFs with the peak frequency in the frequency range of the tremor will be a part of the reconstructed signal. Using this technique, we successfully enhanced the accuracy of our logarithmic model.
Acknowledgements

I am forever indebted to my academic supervisor Prof. Parham Aarabi for his enthusiasm, guidance, patience, and unrelenting support. He has routinely gone beyond his duty to fight my worries, concerns, and anxieties. He has been so supportive since the day I began working in the Mobile Applications Lab. Since then, Parham has supported me, not only by providing research assistantship over almost five years, but also academically and emotionally on the rough road to finishing this thesis.

Parham is one of the smartest people I know. He was and remains my best role model for a scientist, mentor, and teacher. I hope that I can be as lively, enthusiastic, and energetic as Parham and to someday be able to command an audience as well as he can. He has taught me, both consciously and unconsciously, how good experimental research is done. I appreciate all his contributions of time, ideas, and funding to make my Ph.D. experience productive and stimulating.

Also, special thanks to our motivated team from Mount Sinai hospital, St. Michael’s hospital, and Women’s College hospital who have been supportive in every way. My sincere thanks goes to Dr. Bjug Borgundvaag, Dr. Mel Kahan, Dr. Sara Gray, Dr. Shelley McLeod, Sally Carver, Dear Taylor, Taryn Rohringer, and Simon Bromberg.

I am truly grateful to my family for their immeasurable love and care. My hard-working parents have sacrificed their lives for me and my brothers and provided unconditional love and care. They have always encouraged me to explore my potential and pursue my dreams. I love them so much and I would not have made it this far without them. Also to my brothers who have been my best friends all my life, I love them dearly.

Finally, I would like to thank my dear Hossein for his true love, unwavering support and encouragement. He has been the strongest wall to lean on. Thank you for holding my hand and lifting me up in difficult days, and cheering for me all throughout.
## Contents

1 Introduction .................................................. 1
   1.1 Background .............................................. 3
   1.2 Contributions .......................................... 4
   1.3 Outline .................................................. 5

2 Prior Work ................................................... 6
   2.1 Parkinson’s Disease vs. Essential Tremors .............. 6
   2.2 Tremor Recordings from Joint Movements ................. 7
   2.3 Machine Learning Techniques ............................. 7
   2.4 Time-Frequency Analysis of Tremors ...................... 8
   2.5 Empirical Mode Decomposition .......................... 9
   2.6 Tremor Rating Scales ................................... 11

3 Design Decisions ............................................. 13
   3.1 Physical Measurement .................................. 13
      3.1.1 Accelerometer ...................................... 13
      3.1.2 Sampling Rate ...................................... 14
      3.1.3 Sensor Placement .................................. 15
      3.1.4 Device Impact ...................................... 15
      3.1.5 Measurement Duration ............................... 15
      3.1.6 Data Collection .................................... 15

4 Time-Frequency Analysis of Alcohol Withdrawal Tremors .... 17
   4.1 Time-Frequency Analysis ................................. 18
   4.2 Examples of Spectrograms ................................. 18
   4.3 Creation of an iOS Data Collection App .................. 21
      4.3.1 Development of Application ......................... 21
      4.3.2 Data Collection ..................................... 22
4.4 Real vs. Factitious AW Tremors .............................................. 22
4.5 Evaluation of Severity of AW Tremors using a Logarithmic Model .... 24
  4.5.1 Modelling the CIWA-Ar Tremor Rating .................................. 25
  4.5.2 Training the Model .......................................................... 25
4.6 Voluntary vs. Involuntary Tremor Components ................................. 26
4.7 Handedness of Tremors .......................................................... 32
4.8 Summary .................................................................................. 35

5 Performance Evaluation and Clinical Validation ................................. 37
  5.1 Statistical Analysis .................................................................... 37
    5.1.1 P-Value ............................................................................. 38
    5.1.2 Confidence Interval ........................................................... 38
    5.1.3 Root Mean Square Error ..................................................... 38
    5.1.4 Cohen's Kappa Agreement Measure ..................................... 38
  5.2 Analysis Process ........................................................................ 39
    5.2.1 Frequency Analysis ............................................................ 40
  5.3 Effect of EMD in Improving the Logarithmic Model ......................... 41
    5.3.1 Experimental Results .......................................................... 42
    5.3.2 Example of Removal of Voluntary Movements ....................... 42
    5.3.3 Evaluation of Performance of EMD depending on Tremor Severity 43
      5.3.3.1 Example of 'No Observable' Tremor: ................................. 45
      5.3.3.2 Example of 'Mild' Tremor: .............................................. 48
      5.3.3.3 Example of 'Severe' Tremor: .......................................... 50
  5.4 Effect of EMD on Classification of Tremors .................................... 52
  5.5 Analysis of the Logarithmic Model .............................................. 53
    5.5.1 Testing the Model .............................................................. 53
    5.5.2 Error of the Proposed Logarithmic Model ............................... 54
    5.5.3 Inter-rater Variability ......................................................... 54
  5.6 Combining Severity Scores for Both Hands ................................... 58
  5.7 Accelerometer Signal Variability in Different Devices ....................... 60
    5.7.1 Model Correction ............................................................... 63
  5.8 Clinical Validation of the Logarithmic Model .................................. 65
    5.8.1 Study Design and Setting .................................................. 65
    5.8.2 Selection of Participants .................................................... 66
    5.8.3 Methods and Measurements .............................................. 66
    5.8.4 Clinical Validation of the Logarithmic Model .......................... 66
5.8.5 Outcomes ............................................. 67
5.8.6 Results ................................................. 67
   5.8.6.1 Demographics .................................. 67
   5.8.6.2 Derivation and Validation of the Logarithmic Model ... 68
   5.8.6.3 Clinical Validation of the Tremor Model ............ 69
   5.8.6.4 Comparison of Tremor Score based on the Consensus Severity Rating and Clinical Nursing Assessments ............ 70
5.8.7 Importance of Appropriate Tremor Severity Estimation .......... 71
5.8.8 Discussion ........................................... 73
5.8.9 Limitations ........................................... 75
5.9 Summary .............................................. 75

6 Conclusion ............................................. 77

Bibliography ............................................ 78

Appendices .............................................. 88

A Medical Protocols ..................................... 89
List of Tables

1.1 The generated list of all potentially relevant domains and items to be considered for inclusion in a measurement scale to evaluate the severity of AWS in the ED. .......................................................... 2

4.1 RMSE table based on limiting tremor activity to different frequency ranges. 29

5.1 Agreement table. ................................................................. 39

5.2 The effect of EMD noise removal on severity estimation of different groups of tremors. ......................................................... 46

5.3 Inter-rater variability. .............................................................. 57

5.4 Average RMSE for different groups of raters in different tremor rating ranges. 57

5.5 Left-Right tremor data combination methods and their accuracy in estimating CIWA-Ar tremor score. ........................................ 58

5.6 Average rating and rating standard deviation in different iOS devices for the weak vibration platform. ........................................ 62

5.7 Average rating and rating standard deviation in different iOS devices for the strong vibration platform. .................................... 62

5.8 Characteristics and emergency department demographics of included patients. ................................................................. 68

5.9 The logarithmic scores compared to the consensus expert tremor scores during the derivation of model. .................................... 69

5.10 The logarithmic scores compared to the consensus expert tremor scores during the validation of the model. ............................. 70

5.11 RN scores compared to the consensus expert tremor scores. .......... 70

5.12 Comparison of the ED length of stay and diazepam administration for cases of appropriate vs. inappropriate CIWA-Ar protocol. ............... 72

5.13 Characteristics and emergency department demographics of included patients. ............................................................... 72
List of Figures

2.1 Sifting process and envelopes [55] ................................. 10

3.1 Outline of the physical data collection from accelerometer. .... 14
3.2 Example of accelerometer data collection from an AW patient. 16

4.1 A clearly identifiable example of a real tremor .................. 19
4.2 A clearly identifiable example of a factitious tremor .......... 20
4.3 A misclassified example of a real tremor ....................... 20
4.4 A misclassified example of a factitious tremor ............... 21
4.5 Energy of the tremor signal in [5, 15] Hz vs. mean peak frequency of the tremor signals .......... 23
4.6 RMSEs for different frequency ranges based on the training set. 26
4.7 Relation between consensus CIWA-Ar tremor score and energy of the tremor signal in [5, 15] Hz based on the training set ($S = 4.96 \log(1 + 0.067E)$) .... 27
4.8 Original tremor recording with noise and voluntary movements. 30
4.9 Decomposition of the original tremor recording into its IMF components ................. 30
4.10 PSD of the IMF components of the tremor recording .......... 31
4.11 Filtered tremor recording based on reconstruction using only the first three IMFs .......... 32
4.12 Relation between left hand and right hand tremor frequencies. 33
4.13 Relation between left hand and right hand CIWA-Ar tremor ratings by an expert panel of three senior physicians. Please note that there are a total of 61 points in the above graph, with most points overlapping at the same location. 34
4.14 Relation between estimated left hand and right hand CIWA-Ar tremor ratings using the logarithmic model described in Section 4.5. ... 34
4.15 Outline of tremor assessments used in different stages of the analysis. Two patients had assessments from both of their hands. In each case, one assessment was categorized as Real Characteristic Tremor (mild) and the other was categorized as Real Characteristic Tremor (moderate/severe).
5.24 CONSORT diagram of patients in the derivation and internal validation phase. .................................................. 76
A.1 SHOT protocol [16] .................................................. 89
A.2 CIWA-Ar protocol [16] .................................................. 90
Acronyms

3D  3-Dimensional.
ANS  Autonomic Nervous System.
argmax  Argument of Maximum.
AVG  Average.
AW  Alcohol Withdrawal.
AWS  Alcohol Withdrawal Syndrome.
BP  Blood Pressure.
CI  Confidence Interval.
CIWA-Ar  Clinical Institute Withdrawal Assessment for Alcohol, revised.
CNS  Central Nervous System.
CONSORT  Consolidated Standards of Reporting Trials.
CTAS  Canadian Triage and Acuity Scale.
dB  DeciBel.
DSP  Digital Signal Processing.
DTFT  Discrete Time Fourier Transform.
DWT  Discrete Wavelet Transform.
ED  Emergency Department.
EMD  Empirical Mode Decomposition.
ERB  Ethics Review Board.
ET  Essential Tremor.
ETOH  Alcohol.
FT  Fourier Transform.
GI  Gastrointestinal.
HF  High Frequency.
HR  Heart Rate.
IMF  Intrinsic Mode Function.
IQR  Interquartile Range.
IRB  Institutional Review Board.
IRLS  Iteratively Re-weighted Least Square.
LERP  Linear Interpolation.
LF  Low Frequency.
LOS  Length of Stay.
LS  Least Square.
MD  Medicinae Doctor.
ML  Machine Learning.
NN  Neural Network.
PD  Parkinson’s disease.
PSD  Power Spectral Density.
QE  Quadratic Error.
RMSE  Root Mean Square Error.
RN  Registered Nurse.

SD  Squared Difference.

STD  Standard Deviation.

STFT  Short-Time Fourier Transform.

SVM  Support Vector Machine.

TF  Time Frequency.

VAR  Variance of a Random Variable.
Chapter 1

Introduction

Alcohol consumption is the most commonly used mind-altering medication worldwide [1]. Between 10% to 30% of all Emergency Department (ED) visits are related to alcohol consumption [2], [3]. Alcohol Withdrawal Syndrome (AWS) develops in 13-71% of individuals who suddenly stop or reduce their drinking after prolonged and heavy consumption. AWS may be fatal if not treated appropriately [1], [5], [6]. Pharmacologically, alcohol acts as a Central Nervous System (CNS) depressant, and reduced consumption results in the characteristic neuronal hyper-excitability of AWS. Between 5% to 15% of individuals who develop AWS will subsequently develop the major symptoms of withdrawal, including seizures, hallucinations and delirium [7].

Benzodiazepines, the treatment of choice for AWS, are effective in preventing seizures or delirium and treating symptoms [2], [5], [9], [10], [11], [12], [13]. Benzodiazepines pharmacologically mimic alcohol, and their dosing requires careful consideration to alleviate symptoms without excessive sedation [4], [14], [15]. This class of medication is also commonly abused recreationally; patients present to the ED claiming to be in alcohol withdrawal simply to obtain benzodiazepines. Being able to accurately identify and assess the severity of AWS is, therefore, an essential clinical skill.

AWS is best managed using a symptom-driven approach, in which patients are assessed hourly using a standardized tool, and aggressively treated with benzodiazepines based on symptom severity [2], [9], [13], [16], [17]. The Clinical Institute Withdrawal Assessment for Alcohol, revised (CIWA-Ar) [17] is the most commonly used tool for AWS assessment. It covers 10 different domains (nausea/vomiting, tremor, diaphoresis, anxiety, agitation, tactile, auditory and visual hallucinations, headache and orientation). Though clinically validated, it is time consuming for the administer (affecting ED usability), and subjective (many domains are impossible to measure reliably [16], [18]). Table 1.1 summarizes
the list of all potentially relevant domains and items to be considered for inclusion in a measurement scale to evaluate the severity of AWS in emergency departments.

Table 1.1: The generated list of all potentially relevant domains and items to be considered for inclusion in a measurement scale to evaluate the severity of AWS in the ED.

<table>
<thead>
<tr>
<th>DOMAIN</th>
<th>ITEM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Affect</td>
<td>Anxiety</td>
</tr>
<tr>
<td></td>
<td>Restlessness</td>
</tr>
<tr>
<td></td>
<td>Agitation</td>
</tr>
<tr>
<td></td>
<td>Depression</td>
</tr>
<tr>
<td></td>
<td>Craving for Alcohol</td>
</tr>
<tr>
<td></td>
<td>Asthenia</td>
</tr>
<tr>
<td>GI Disturbance</td>
<td>Anorexia</td>
</tr>
<tr>
<td></td>
<td>Nausea</td>
</tr>
<tr>
<td></td>
<td>Vomiting</td>
</tr>
<tr>
<td></td>
<td>GI Disturbance</td>
</tr>
<tr>
<td></td>
<td>Abdominal Pain</td>
</tr>
<tr>
<td>ANS Disturbance</td>
<td>Temperature</td>
</tr>
<tr>
<td></td>
<td>Diaphoresis</td>
</tr>
<tr>
<td></td>
<td>Flushing</td>
</tr>
<tr>
<td></td>
<td>Tachycardia</td>
</tr>
<tr>
<td></td>
<td>Palpitations</td>
</tr>
<tr>
<td></td>
<td>Hypertension</td>
</tr>
<tr>
<td></td>
<td>Headache</td>
</tr>
<tr>
<td>Neurological Disturbance</td>
<td>Tremor</td>
</tr>
<tr>
<td></td>
<td>Impaired Co-ordination</td>
</tr>
<tr>
<td></td>
<td>Altered Consciousness</td>
</tr>
<tr>
<td></td>
<td>Concentration</td>
</tr>
<tr>
<td></td>
<td>Dizziness</td>
</tr>
<tr>
<td>Sleep Disturbance</td>
<td>Insomnia</td>
</tr>
</tbody>
</table>
Chapter 1. Introduction

<table>
<thead>
<tr>
<th>Psychotic Features</th>
<th>Other Sleep Problems</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Hallucinations</td>
</tr>
<tr>
<td></td>
<td>Visual Disturbance</td>
</tr>
<tr>
<td></td>
<td>Auditory Disturbance</td>
</tr>
<tr>
<td></td>
<td>Tactile Disturbance</td>
</tr>
<tr>
<td></td>
<td>Delusions</td>
</tr>
<tr>
<td>Other</td>
<td>Seizures</td>
</tr>
</tbody>
</table>

Among the clinical signs of withdrawal, tremor is the most prevalent sign in all withdrawal severity scales [19]. Even so, the ability to quantify alcohol withdrawal tremor severity is highly variable and dependent on clinical experience. Furthermore, the scale provides no description of the characteristic features of a typical AWS tremor to assist inexperienced clinicians with its application. A standardized method of assessing the tremor of AWS would greatly improve the accuracy and reliability of the CIWA-Ar, and may allow the development of a shorter, more useful ED tool for assessing AWS severity [16], [20], [21], [22]. Medical protocols for CIWA-Ar and SHOT, the short version of CIWA-Ar, are attached in appendix A.

1.1 Background

There have been numerous studies conducted on Parkinson’s disease and essential tremors previously [23]. Researchers have investigated the peak frequency as well as the frequency range of Parkinson’s and essential tremors using time-frequency techniques. There are further studies on how to differentiate a Parkinson’s tremor from an essential tremor [24], [25]. In each of these studies different signal acquisition techniques, acquisition location, and different types of sensors have been used.

In addition to the signal processing techniques, some have studied the use of Machine Learning (ML) algorithms to differentiate a Parkinson’s disease tremor from an essential tremor and classify them. Some have also used Support Vector Machines (SVM) [24] and others have applied Neural Network (NN) to learn the features of each class of tremors [26].
Aside from classifying the tremors, another set of analysis have been conducted in order to quantify or detect essential tremors and Parkinson’s tremors. Researchers have applied adaptive Fourier modelling [27] to identify essential tremors. In [28], Power Spectral Density (PSD) of the recorded signal from patients with Parkinson’s disease has been computed. Furthermore, researchers in [28] have introduced a technique to record and quantify the tremor signal from the index finger of patients with an essential tremor during their daily activity to help patients with monitoring their tremor.

In 1996, the idea of using the writing or drawings of patients with an essential tremor to quantify the severity of their tremors was introduced [29]. Two years later, [30] studied a new technique involving drawing a spiral by a patient to quantify the tremor severity. Researchers have investigated this technique further using different Machine Learning (ML) algorithms [31], [32], [33].

1.2 Contributions

As noted in the background section, most of the studies thus far have focused on Parkinson’s disease tremors or essential tremors. Although being frequently encountered in emergency departments, there has not been much work done to improve the care of patients with AW. Therefore, in this thesis we have set out the following:

- Studying features of AW tremor signal such as frequency and energy [34].

- Quantifying severity of AW tremors using a logarithmic model and validating the experimental model on a set of patients [35].

- Examining the difference in the intensity and frequency of the tremor signals in both hands of patients with different handedness [36].

- Evaluating the consistency of different accelerometers in recording tremor signals in three dimensions in order to quantify the severity of the AWS tremor.

- Using EMD algorithm to remove the noise and voluntary component of hand tremor from the AW tremors recordings. This can enable one to quantify the tremor severity better and improve the logarithmic model [37].
• Validating the study in a clinical setting by deploying an iOS application in emergency departments. Each of the patients were administered by a nurse or a physician and videotaped simultaneously. Furthermore, the accuracy of the model was compared against the consensus severity score given by a panel of three expert physicians after watching those videos [38], [39], [40], [41], [42].

As a result of this project a signal acquisition platform has been established for emergency departments in order to quantify the severity of the tremor signals in patients with AW.

1.3 Outline

The structure of this thesis is comprised of six chapters. The first chapter is the introduction and a brief description of its main contributions. In the second chapter we review previous studies, which have been done for different types of tremors and introduce signal processing techniques used as a part of the analysis. The third chapter contains the design fundamentals. In Chapter four the main elements of time-frequency analysis on AW tremors are explained, as well as the differences between fake and real tremors, the underlying logarithmic nature of AW tremors, effect of the EMD algorithm to remove the noise and voluntary hand movement, and handedness of tremors.

In Chapter five the performance of our model is evaluated on a set of patients (separate from training set). The effectiveness and accuracy of the model in a clinical setting is also validated.

The last chapter summarizes all the findings throughout the thesis.
Chapter 2

Prior Work

A tremor is an involuntary oscillatory movement of a body part. The most common types of tremors are Physiologic tremors, essential tremors, Parkinson’s disease tremor, and alcohol withdrawal tremor. There are two types of questions that might arise while studying tremor signals. First, how can one classify tremors based on studying their signals, and second, how can one quantify the severity of the tremor and treat the patients according to appropriate medical protocols.

In this chapter we evaluate different studies being conducted in this field, either in classifying different types of tremors or in measuring the intensity, frequency, and amplitude of tremors [43].

2.1 Parkinson’s Disease vs. Essential Tremors

Essential Tremor (ET) is the most common form of tremor and most common movement disorder in general [44]. Parkinson’s Disease (PD) tremor is the second most commonly diagnosed tremor. Tremor is one of the key features of Parkinson’s Disease, yet 25% of patients with PD will never develop a tremor.

Time-Frequency analysis and visualization is one of the classic methods in understanding and evaluating different tremor signals. In one study, researchers collected the tremor signal simultaneously from different muscles of patients with PD and ET [45]. They then calculated the time-frequency distribution of each tremor as a filtered Wigner distribution and used this to distinct oscillator systems underlying tremors in different limbs. Furthermore, it has been established that in patients with PD and ET the instantaneous tremor frequency can change by a fraction of 1 Hz over a period of seconds. Another key finding was that by revealing frequency dissociation among physically contracting mus-
cle groups researchers were able to distinguish psychogenic tremors from non-psychogenic tremors.

The prominent defining tremor frequency in PD in resting condition is within 4 to 6 Hz. On the other hand, the peak frequency of the essential tremor is typically 4 to 7 Hz \[46\]. Due to these similarities between the two types of tremors, about 25% of tremors are misdiagnosed in the emergency department. In 2012, researchers used the changes in the tremors to extract statistically significant information from the tremor signal and effectively differentiate between PD and ET conditions \[17\]. Discrete Wavelet Transform (DWT) was used first to extract the tremor component with a frequency in the range of 4 to 7 Hz from the other components of the recording. Support vector machines were then utilized to extract the key features of PD and ET \[24\], \[48\].

Another study demonstrated that researchers can distinguish PD and ET from the fluctuations in the tremor signal while subjects were in the resting condition. Based on their experimentation on 52 subjects, it was concluded that the ratio of fluctuations of resting to a kinetic task is a sensitive feature to discriminate these two types of tremors with high accuracy \[25\].

2.2 Tremor Recordings from Joint Movements

There are different sensors and devices used to record tremor signals \[49\]. In one study, researchers measured the tremor at the body joint of patients in the affected limb. Knowing the differences between the tremors generated by different muscles could help in finding local botulinum toxin injection. Using multiple sensors on different locations of the body and considering two degrees of freedom for each, researchers were able to find the muscle groups most involved in the developed tremor. This achievement can potentially help in characterizing different tremors \[50\].

2.3 Machine Learning Techniques

Different machine learning techniques have been utilized to classify tremors and also investigate the severity of tremors. Among those, is the use of the support vector machine to extract features of different tremors \[24\] and effectively classify PD and ET. Secondly, other studies used neural networks to classify tremors \[51\], \[52\]. The NN took linear prediction coefficients, wavelet transform based variance and entropy values, higher-order cumulants, and power ratio as key features in training and testing stages. It appeared that
wavelet transform based variance, entropy values, and higher-order cumulants were not
good features for classification purposes, whereas power ratio appeared to be a significant
feature for all tremor types.

## 2.4 Time-Frequency Analysis of Tremors

The Fourier Transform (FT), \( X(\omega) \), of a signal \( x(t) \) is defined as:

\[
X(\omega) = \int_{-\infty}^{\infty} x(t)e^{-j\omega t}dt
\]  

(2.1)

where \( t \) and \( \omega \) are the time and frequency parameters, respectively. It defines the spec-
trum of \( x(t) \), which consists of components at all frequencies over the range of which it
is nonzero.

Historically, Fourier spectrum analysis has provided a general method for examin-
ing the global energy-frequency distribution. Fourier analysis has dominated the data
analysis efforts since its introduction, because of its power and simplicity. The Fourier
transform belongs to the class of orthogonal transformations that uses fixed harmonic
basis functions. The Fourier transform result can be shown as a decomposition of the
initial signal into harmonic functions, with fixed frequencies and amplitudes.

For many signals, Fourier analysis is useful because the signal’s frequency content
is important. But Fourier analysis has a serious drawback for information loss while
transforming the signal to its frequency domain. It is only valid under extremely general
conditions, (i.e. the system must be linear, and the data must be strictly periodic or
stationary) otherwise the resulting spectrum will be physically unreliable.

Dannis Gabor in 1946 adapted the Fourier transform to analyze only a small set of
signals at a time [53]. It is called Short-Time Fourier Transform (STFT). The STFT is
obtained from the usual FT by multiplying time domain signal \( x(t) \) by an appropriate
sliding time window \( w(t) \). Thus, instead of the usual FT expression 2.1 gets a time-
frequency expression of the form:

\[
X(\tau,\omega) = \int_{-\infty}^{\infty} x(t)w(t-\tau)e^{-j\omega t}dt
\]  

(2.2)

where \( w(t) \) is the sliding time window applied to the signal.

The information STFT provides has limited precision, which is determined by the size
of the window.
2.5 Empirical Mode Decomposition

Empirical Mode Decomposition (EMD) \cite{54} is a method of breaking down the signal without leaving the time domain. This algorithm is based on the decomposition of the original signal into a collection of Intrinsic Mode Functions (IMFs) using a numerical sifting process.

IMFs must fulfill two conditions:

i) the number of maxima and the number of zero crossings must be equal or different at most by one; and

ii) the mean value between the upper and lower envelopes is zero everywhere.

The sifting process can be separated into following steps:

For a signal $x(t)$, let $m_1$ be the mean of its upper and lower envelopes as determined from a cubic-spline interpolation of local maxima and minima. The locality is determined by an arbitrary parameter; the calculation time and the effectiveness of the EMD depends greatly on such a parameter.

The first component $h_1$ is computed as:

$$ h_1 = x(t) - m_1 \quad (2.3) $$

- In the second sifting process, $h_1$ is treated as the data, and $m_{11}$ is the mean of $h_1$’s upper and lower envelopes:

$$ h_{11} = h_1 - m_{11} \quad (2.4) $$

- This sifting procedure is repeated $k$ times, until $h_{1k}$ is an IMF, that is:

$$ h_{1(k-1)} - m_{1k} = h_{1k} \quad (2.5) $$

- Then, it is designated as $c_1=h_{1k}$, the first IMF component from the data, which contains the shortest period component of the signal. We separate the first IMF from the rest of the data:

$$ x(t) - c_1 = r_1 \quad (2.6) $$

- The procedure will be repeated on $r_j : r_1 - c_2 = r_2, ..., r_{n-1} - c_n = r_n$. 
Thus the original signal $x(t)$ can be expressed as:

$$x(t) = \sum_{j=1}^{n} c_j(t) + r_n(t)$$

(2.7)

- $c_j(t)$ is an IMF where $j$ represents the number of corresponding IMF and $r_n(t)$ is residue.

The sifting process, the maximum, minimum, and the mean envelopes are shown in Figure 2.1 [55].

![Figure 2.1: Sifting process and envelopes](image)

The stoppage criteria determines the number of sifting steps to produce an IMF. Two different stoppage criteria are being used traditionally:
Chapter 2. Prior Work

1. This criterion is proposed by Huang et al. [56] and defined as a sum of the squared differences, SD,

\[
SD_k = \sum_{t=0}^{T} |h_{k-1}(t) - h_k(t)|^2 \quad \sum_{t=0}^{T} h_{k-1}^2(t)
\]

Then, the sifting process will stop when SD is smaller than a pre-given value.

2. A second criterion is based on a number called the S-number. The S-number is defined as the number of consecutive siftings when the numbers of zero-crossings and extrema are equal or at most differ by one. Normally, an S-number is pre-selected. The sifting process will stop only if for S consecutive iterations the numbers of zero-crossings and extrema stay the same, and are equal or at most differ by one.

The EMD decomposes signals into narrow-band components with decreasing frequency. The decomposition is complete, almost orthogonal, local and adaptive. All IMFs form a complete and “nearly” orthogonal basis for the original signal. The basis directly comes from the signal, which guarantees the inherent characteristic of the signal and avoids the diffusion and leakage of signal energy.

2.6 Tremor Rating Scales

Tremor rating scales for different types of tremors are used in the clinical assessment of tremors. In the literature of Parkinson’s disease tremor and essential tremor, different models for estimating tremor rating is proposed.

According to Weber’s law of Phychophysics, the smallest change in the tremor amplitude is correlated with the initial tremor amplitude:

\[
\Delta T = K \times T_1
\]

where \(T_1\) is the initial amplitude of the tremor and \(K\) is Weber’s constant [57].

Weber’s law suggests that any perception of the tremor severity can be measured in a logarithmic scale. Specifically, Deuschl, Elble, and their team described and validated logarithmic models that can relate the clinical rating score to the amplitude of the tremor for patients with essential tremor [58], [59], [60]. Their work established that based on different datasets collected from patients with an essential tremor in different conditions,
a logarithmic relationship can be observed. However, there is no recommendation as to what is the exact model that can predict the clinical tremor rating based on the amplitude.

Furthermore, in another study the relationship between the readings from an accelerometer and gyroscope during daily activity of patients with essential tremor and the average clinician rating of the tremor was investigated and a model of form Equation 2.10 was proposed [23].

\[ R = b_0 + B_a \times P_a + B_g \times P_g \]  

(2.10)

where \( R \) is the average clinician rating, \( P_a \) and \( P_g \) are the peak powers from readings of the accelerometer and gyroscope, and \( b_0, B_a \) and \( B_g \) are the regression coefficients.
Chapter 3

Design Decisions

The structure of this thesis involves design-specific decisions. Outlined in this chapter, the decisions regarding the physical measurement of tremors, including sensor type, sampling rate, sensor placement, device impact, and data collection steps will be described.

3.1 Physical Measurement

In this section, the design of the physical measurement of the tremor activity is described. The tremor signal was collected using an accelerometer. Deciding the type of accelerometer, the sampling frequency, and the acquisition location requires careful consideration.

The process of the physical measurement of the tremor using an accelerometer is described in Figure 3.1. After admitting the patient in the ED and enrolling him/her on CIWA-Ar protocol, they were recorded using accelerometer for 20 seconds from the hand. At the same time our clinical staff videotaped the data collection process for the panel of expert physicians to review the videos and rate the severity of tremors.

3.1.1 Accelerometer

An accelerometer was used in a mobile platform. A mobile platform was preferred because a user-interface was required to provide the severity estimation of the tremor in realtime. Among all mobile platforms, an Apple iPod Touch was chosen due to the reasonable price and feasibility of using it in the emergency department. The accelerometer in Apple iPod Touch 5th generation is “STMicroelectronics LIS331DLH” accelerometer
Figure 3.1: Outline of the physical data collection from accelerometer.

[61], which measures acceleration in units of $g$, and gives a maximum frequency of 100 Hz. The data sheet for this sensor is available in [62].

### 3.1.2 Sampling Rate

With the Apple iPod Touch 5th generation, one can have a sampling rate between 0.1 Hz up to 100 Hz. A sampling rate of 65 Hz was selected for two main reasons:

1) Sampling above the Nyquist rate.
2) Compatibility with historical data.
3.1.3 Sensor Placement

Patients were holding the iPod in their outstretched palm while seated. This approach was in line with the way physicians rated the severity of AW tremors. They looked at the patients’ palms to give their severity estimation.

3.1.4 Device Impact

A damping effect is expected on the tremor when patients are holding the device in their hand. This is mostly due to the size and weight of the iPod. The Apple iPod Touch 5th generation is $123.4 \times 58.6 \times 6.1\,mm$ with a total weight of 88 grams [63]. The weight and size of the device, as well as patient attempts to control the tremor to keep the device from falling could result in a bias in the tremor recordings.

3.1.5 Measurement Duration

Accelerometer data was recorded from both hands of each patient for 20 seconds. This is due to the limited duration that the patient can hold the device without dropping it. Also, based on physicians’ experience and suggestion, 20 seconds was chosen to reliably record the accelerometer data without tiring the patient.

3.1.6 Data Collection

Accelerometer data is collected from the hands of patients in alcohol withdrawal for 20 seconds with a sampling rate of 65 Hz. For each assessment, there were accelerations in three dimensions for 20 seconds. For most patients during the training step, data from both hands was collected. Figure 3.2 demonstrates the data collection from an AW patient.
Furthermore, medical staff in the emergency department are videotaping hand movement of the patient at the time of the assessment in order for the expert physicians to review and rate their severity. The consensus severity score from the panel of three expert physicians was used as the ground truth throughout the thesis.
Chapter 4

Time-Frequency Analysis of Alcohol Withdrawal Tremors

Time-Frequency analysis comprises of a study of signals in two domains simultaneously, time and frequency. This technique represents the signal in a 2-dimensional space of frequency and time. While studying AW tremors, it is easier to visualize frequency and time together, since these two domains are tightly connected. Another reason to use time-frequency analysis in analyzing tremor signals is the fact that our tremor recordings are short in duration and can change substantially over time and studying these changes can help us to quantify the severity of AW tremors.

Time-Frequency analysis of a signal is simple, fast, and enables one to see different biomedical phenomena, such as fluctuations of the peak frequency, range of the dominant energy, and variability and spread of the frequency range of the signal.

The first step of analysis is to examine the effect of frequency and band-limited energy in measuring the severity of AW tremors. The second step is finding the relationship between the band-limited energy of tremors and the severity of them and how one can define this relationship. Is this relationship strong, moderate or weak? Is it logarithmic, linear or exponential? Is it increasing or decreasing?

Finally, the relationship between the severity of the tremors assessed by physicians and the energy of the tremor signal, whether that differs from one hand to the other is explored.
4.1 Time-Frequency Analysis

Often with biologically-derived periodic signals such as speech, heart signals, or body motion, a time-frequency view can provide unique insight [64]. Spectrograms are used in order to better understand the nature and type of tremors. Short-time Fourier transform with a Hanning window of 4 seconds and 90% overlap were generated. No major change was noted by using other window functions. The peak frequency (in the 0 - 15 Hz range) of each window was used for tremor detection analysis as shown below:

$$V(n) = \arg\max_k |X_n(k)|$$

(4.1)

where $X_n(k)$ is the STFT of window $n$ at frequency index $k$.

We define the mean peak frequency $\mu$ as:

$$\mu = \frac{1}{N} \sum_{n=1}^{N} V(n)$$

(4.2)

and the Quadratic Error (QE) (relative to the mean) as:

$$\sigma = \sqrt{\frac{1}{N} \sum_{n=1}^{N} (V(n) - \mu)^2}$$

(4.3)

with $N$ being the total number of windows during 20 seconds.

The band-limited energy of the tremor signal in the intervals $[f_0, 15]$ Hz was computed. The energy of the tremor signal in the $[f_0, 15]$ Hz range is defined as the sum of the squared time-frequency blocks in this range as shown bellow:

$$E = \sum_n \sum_k |X_n(k)|^2$$

(4.4)

where $X_n(k)$ is the STFT of the window $n$ at frequency index $k$ in the interval $[f_0, 15]$ Hz.

4.2 Examples of Spectrograms

Figures 4.1 and 4.2 illustrate examples of spectrograms from two tremor recordings. Figure 4.1 shows a spectrogram from a patient with a real AW tremor. The dominant and consistent frequency for this assessment is 9.7 Hz with a QE of 0.3 Hz. Figure 4.2
illustrates a factitious tremor without a consistent frequency peak. In this case, the mean peak frequency is 6.3 Hz. This assessment was identified as a factitious tremor based on our empirically determined cut-off of 7 Hz that will be described in the next section. One other reasoning for this classification could be based on inconsistent tremor peak frequency. As is evident, there is no visible and identifiable tremor component in the first 10 seconds of the recording.

Figures 4.3 and 4.4 show misclassified (based on expert physician rating) examples from the data. Figure 4.3 shows a real tremor with a mean peak frequency of 5.6 Hz. This would be incorrectly classified as a factitious tremor. Figure 4.4 illustrates a factitious tremor with a mean peak frequency of 8.3 Hz. This would be incorrectly classified as a real tremor. However, the real tremor in Figure 4.3 has a more consistent dominant frequency compared to the factitious tremor shown in Figure 4.4.

Figure 4.1: A clearly identifiable example of a real tremor.
Figure 4.2: A clearly identifiable example of a factitious tremor.

Figure 4.3: A misclassified example of a real tremor.
4.3 Creation of an iOS Data Collection App

In collaboration with Simon Bromberg, an iOS application was created for the use of the data collection in the emergency department. This application used the built-in accelerometer of a handheld iOS device to measure the magnitude of accelerations in three dimensions. Data was collected over a 20-second window with a sampling rate of 65 Hz. The base analysis was performed in Matlab and the final algorithm was implemented in an iOS platform. The entire data collection in this thesis was conducted using Apple iPod Touch 4th and 5th generation and iPhone 5s.

4.3.1 Development of Application

The described application is the first CIWA-Ar application to assist clinicians with evaluating alcohol withdrawal severity. The novel component of this application is the use of the built-in accelerometer in a handheld iOS device to objectively quantify and characterize the tremor component of alcohol withdrawal syndrome. Our app is able to reliably quantify tremor based on our logarithmic model, which will be described in next
section. The accuracy of the model is at least as good as experts in providing a severity rating for tremors. Results will be discussed in Chapter 5 [35], [39].

4.3.2 Data Collection

This research was conducted in two large, urban, university teaching hospital emergency departments starting in July 2013, and was approved by the Ethics Review Board (ERB) at each hospital. Patients presenting to the emergency department in alcohol withdrawal were approached by a research assistant and provided informed consent for the research. For each assessment, patients were instructed to hold the iOS device in their right hand with their arm fully extended for 20 seconds. Throughout the course of the study, it was observed that the tremor would vary between left and right hands, and the protocol was subsequently modified to include tremor assessments from both hands.

Each assessment was videotaped, and retrospectively viewed and rated by a panel of 3 expert clinicians using the 8-point CIWA-Ar scale (0 = no tremor, 7 = severe tremor). For patients with recordings from both left and right hands, each hand was rated separately. For analysis purposes, expert clinicians also subdivided assessments into five categories:

1) Real Characteristic Tremor (mild);
2) Real Characteristic Tremor (moderate/severe);
3) No Observable Tremor;
4) Atypical Tremor (patient); and
5) Factitious Tremor (nurse).

4.4 Real vs. Factitious AW Tremors

The spectrograms for 137 assessments from real patients were examined and the energy of the tremor signal vs. mean peak frequency compared (Figure 4.5). As demonstrated by this figure, the mean peak frequency could be a useful indicator to differentiate real from factitious tremors. The mean peak tremor frequency for patients with actual observable AW tremor was 7.01 Hz.

In comparison, the mean peak frequency was 6.53 Hz for nurses mimicking the AW tremor and 5.11 Hz for patients with atypical tremors. A variety of other features can be observed, including energy consistency and change of acceleration angle; however, the study focused only on peak frequency.
Chapter 4. Time-Frequency Analysis of Alcohol Withdrawal Tremors

Figure 4.5: Energy of the tremor signal in [5, 15] Hz vs. mean peak frequency of the tremor signals.

Although not conclusive, this mean peak frequency difference suggests that there may be a potential in using spectral information for inferring tremor validity.

With respect to differentiating real versus factitious tremors, the goal was to provide an initial answer as to whether classification is possible based on the electronic tremor recordings. 75% of the recordings from patients with AW syndrome had a mean peak tremor frequency higher than 7 Hz, whereas only 17% of nurses mimicking AW tremor had a tremor with a frequency in this range, suggesting that tremor above 7 Hz could be a potential differentiator of real versus factitious AW tremors. Notably, the selection of the range of mean peak frequency was visually done based on evaluating the spectrograms of patients and nurses mimicking AW tremors. If the peak frequency cut-off was changed to 7.1 Hz, 50% of these potentially factitious tremors would have been captured in the low frequency range without affecting the percentage of ‘real’ tremors falling above the cut-off. More data from patients who are potentially ‘faking’ an AW tremor are needed to evaluate if there is a meaningful difference between a cut-off of 7 versus 7.1 Hz.

It is truly impossible to ascertain whether tremors classified as ‘real’ were in fact legitimate AW tremors; likewise, it is possible that patients’ tremors classified as ‘suspicious/factitious’ were in fact legitimate. The experienced observers did their best to identify potentially factitious tremors based on atypical or uncharacteristic tremor features (e.g., movement in only one axis), which, in practice, is all clinicians have to rely
on in order to differentiate genuine from factitious tremors. This emphasizes the importance of this current investigation, which seeks to characterize typical AW tremors, and to identify useful features for distinguishing real from factitious tremors.

It should be noted that this is a very simple, preliminary view of this data with the goal being to characterize the type of tremor recordings that are observed in actual AW patients. Further data analysis, including machine learning or support vector machines, could provide significant improvements in classification.

### 4.5 Evaluation of Severity of AW Tremors using a Logarithmic Model

There were 100 participants in this step of the research, with a total of 137 tremor assessments. 125 tremor assessments were performed on 88 patients with AW symptoms. An additional 12 assessments were performed on 12 nurses deliberately mimicking an AW tremor (factitious). Recordings from nurses and those from patients with atypical tremors (8 assessments, 5 patients) were excluded from our derivation model. Recordings from all other patients (117 recordings, 83 patients) were used to compare the energy of the tremor signals computed using equation 4.4 with CIWA-Ar ratings from a panel of 3 expert clinicians. Figure 4.15 outlines the assessments and patients included at various steps throughout the analysis.

A logarithmic relationship was observed between the band-limited energy and the consensus tremor severity rating. In order to force the curve to pass through the origin when the energy of the tremor signal is 0, we fitted a logarithmic model of the form:

\[
S = a \cdot \log(1 + b \cdot E)
\]  

(4.5)

where \( S \) is the estimated tremor severity rating, \( E \) is the band-limited energy, and \( a \) and \( b \) are constants.

The model was fitted using nonlinear regression algorithm implementing Iteratively Re-weighted Least Squares (IRLS) [65]. This algorithm tries to minimize the residuals using Least Square (LS) minimization and robust weights are recalculated based on each observation’s residual from the previous iteration [66]. These weights down-weight outliers so that their influence on the fit is minimized.
4.5.1 Modelling the CIWA-Ar Tremor Rating

Tremor assessments for patients included in the analysis were subdivided into two separate sets: a training set and a testing set. The training set (84 assessments, 50 patients) was used to derive a logarithmic model, which was then applied to a separate group of 33 assessments from 33 patients to test the model’s ability to accurately rate the tremor severity. Furthermore, both groups of patient data were used to determine the performance of the model in a combined analysis. The clinical validation of the model on a separate set of patients is described in Chapter 5.

4.5.2 Training the Model

84 tremor recordings from 50 patients were used to compute the band-limited energy of the tremor signal in the intervals \([f_0, 15]\) Hz. Of these 84 recordings, 59 were categorized as real, characteristic, mild tremors, 15 were categorized as real, characteristic, moderate/severe tremors, and 10 were categorized as no observable tremor. The CIWA-Ar tremor severity score assigned by expert clinicians and the energy of the signal in each of the frequency ranges described by equation 4.4 were compared. For each frequency range, we calculated the Root Mean Square Error (RMSE) between each score estimated by the logarithmic model of the form described in equation 4.5 and the score provided from our expert panel. The lowest RMSE demonstrates the best frequency range to look at for estimating the CIWA-Ar tremor severity. Figure 4.6 illustrates the RMSE for all described frequency ranges based on this training set.

The minimum RMSE is 0.85 and corresponds to the frequency range of \([5, 15]\) Hz. The line of best logarithmic approximation used to provide electronic estimation of the CIWA-Ar tremor severity is shown in Figure 4.7. The energy shown here is the sum of the squared time-frequency blocks in range of \([5, 15]\) Hz over 20 seconds of recording. Y axis is the CIWA-Ar tremor severity score according to expert consensus as outlined before. The logarithmic model trained on the the training set using non-linear regression algorithm is of the form \(S = 4.96\log(1 + 0.067E)\) with \(S\) and \(E\) being the estimated tremor severity rating and the band-limited energy in the range of \([5, 15]\) Hz, respectively.

As is evident from Figure 4.7, the data itself suggests a logarithmic pattern. It is in fact inline with what we expect from the magnitude of any vibrational phenomena. The two most encountered examples are using the Richter magnitude scale for measuring the magnitude and intensity of an earthquake, and DeciBel (DB) as a measure of sound.
level, both of which are logarithmic scales.

### 4.6 Voluntary vs. Involuntary Tremor Components

In this section, we explore and demonstrate the ability of the Empirical Mode Decomposition (EMD) technique [56] to extract tremor activity from the accelerometer recording. We then explore the effect of the EMD filtering technique on our proposed logarithmic model in Chapter 5.

EMD decomposes the signal into basic components called Intrinsic Mode Functions (IMFs) [67]. Identification of the physical meaning of each IMF in the context of tremor analysis is essential to automatically distinguish tremor from voluntary activities [68], [69], [70], [71].

IMFs must fulfill two conditions:

1) The number of extrema and the number of zero crossings must be equal or different at most by one; and
ii) The mean value between the upper and lower envelopes is zero everywhere.

The original tremor signal is decomposed using the EMD method following these steps [72]:

1. Initialize $h_0(t)$ with the original signal and set $i = 0$.
2. Identify the extrema of the signal, $h_i(t)$.
3. Generate the upper and lower envelopes by interpolation of maxima and minima points developed in previous step.
4. Calculate the mean of two envelopes to determine the local mean value, $m(t)$.
5. Calculate $d(t) = h_i(t) - m(t)$.
6. Test if $d(t)$ becomes a zero-mean signal, then $d(t)$ is considered as the next IMF, $h_{i+1}(t) = d(t)$. Otherwise, replace $h_i(t)$ with $d(t)$ and repeat from step 2.
7. Update the residue series as \( r(t) = r(t) - h_i(t) \) and \( i = i + 1 \). Repeat steps 2 to 7 by sifting the residual signal. The process is stopped when the final residual signal is obtained as a monotonic function.

After the EMD process, the signal is decomposed into a residue and a collection of IMFs. Hence, it can be expressed as:

\[
\sum_{i=1}^{n} h_i(t) + r
\]

where \( n \) is the number of IMFs.

In our implementation of EMD algorithm we used the default stopping criterion proposed in [73]:

- At each point \( \text{mean} - \text{amplitude} < \frac{\text{envelope} - \text{amplitude}}{2} \)
- Mean of the boolean array \( \{ \frac{\text{mean} - \text{amplitude}}{\text{envelope} - \text{amplitude}} > 0.05 \} \) < 0.05
- \(| \# \text{ of zeros} - \# \text{ of extermas} | \) ≤ 1

where \( \text{mean} - \text{amplitude} = \frac{|\text{envelope}_{\text{max}} + \text{envelope}_{\text{min}}|}{2} \) and;
\( \text{envelope} - \text{amplitude} = \frac{|\text{envelope}_{\text{max}} - \text{envelope}_{\text{min}}|}{2} \).

This criteria is set to aim for small fluctuations globally while at the same time allowing for large changes in the signal locally [74].

Figures 4.8 to 4.11 illustrate the application of the EMD algorithm in each step. Figure 4.8 shows the original tremor recording from an AW patient. Figures 4.9 and 4.10 demonstrate the decomposition of the original tremor recording from a patient with AWS into its IMFs and the corresponding power spectral density of the IMF components, respectively.

It is well-established that voluntary tremor activities are in the frequency range of \([0, 2]\) Hz and the tremor component of the signal tends to lie in \([2, 12]\) Hz [75], [76], [77]. Specially, one study demonstrated that the tremor activity is limited to \([3, 8]\) Hz [78].

The accuracy of our logarithmic model estimation was investigated based on the error measure of RMSE when the focus is on the tremor component in different frequency ranges. Table 4.1 summarizes RMSEs for different models built on the training set of 84 recordings from 50 patients. RMSEs in this table are calculated based on the RMSE
Equation 5.1 with $S_i$’s being the estimated tremor severity score based on the logarithmic model relating the band-limited energy of the signal in the interval $[f_l, f_u]$ and the consensus rating from our panel. Equation 4.4 defines the general definition of band-limited energy in this thesis.

Based on this analysis, there is a marginal difference between different frequency ranges as long as the tremor component is not removed. Therefore, the focus is also on frequency range of $[2, 12]$ Hz as suggested from the literature [75], [76], [77].

Table 4.1: RMSE table based on limiting tremor activity to different frequency ranges.

<table>
<thead>
<tr>
<th></th>
<th>$f_l = 1$ Hz</th>
<th>$f_l = 2$ Hz</th>
<th>$f_l = 3$ Hz</th>
<th>$f_l = 4$ Hz</th>
<th>$f_l = 5$ Hz</th>
</tr>
</thead>
<tbody>
<tr>
<td>$f_u = 8$ Hz</td>
<td>0.72</td>
<td>0.72</td>
<td>0.71</td>
<td>0.71</td>
<td>0.78</td>
</tr>
<tr>
<td>$f_u = 9$ Hz</td>
<td>0.75</td>
<td>0.75</td>
<td>0.75</td>
<td>0.75</td>
<td>0.82</td>
</tr>
<tr>
<td>$f_u = 10$ Hz</td>
<td>0.73</td>
<td>0.73</td>
<td>0.73</td>
<td>0.73</td>
<td>0.80</td>
</tr>
<tr>
<td>$f_u = 11$ Hz</td>
<td>0.73</td>
<td>0.73</td>
<td>0.73</td>
<td>0.73</td>
<td>0.79</td>
</tr>
<tr>
<td>$f_u = 12$ Hz</td>
<td>0.73</td>
<td>0.73</td>
<td>0.73</td>
<td>0.73</td>
<td>0.8</td>
</tr>
<tr>
<td>$f_u = 13$ Hz</td>
<td>0.73</td>
<td>0.73</td>
<td>0.73</td>
<td>0.73</td>
<td>0.8</td>
</tr>
<tr>
<td>$f_u = 14$ Hz</td>
<td>0.72</td>
<td>0.72</td>
<td>0.72</td>
<td>0.72</td>
<td>0.77</td>
</tr>
<tr>
<td>$f_u = 15$ Hz</td>
<td>0.72</td>
<td>0.72</td>
<td>0.72</td>
<td>0.72</td>
<td>0.77</td>
</tr>
</tbody>
</table>

where $f_l$ is the lower limit of frequency range and $f_u$ is the upper limit of the range.

In order to categorize each IMF (tremor or voluntary movement), the Power Spectral Density (PSD) of IMFs using Welch’s method was computed [79]. Welch’s method is used for estimating power spectral density of a signal. The PSD estimate is carried out by windowing the signal, in this case a Hanning window, forming the periodogram for each block, and averaging the energy for each block. The averaging of periodograms tends to decrease the variance of the estimation relative to a single periodogram estimate.
Figure 4.8: Original tremor recording with noise and voluntary movements.

Figure 4.9: Decomposition of the original tremor recording into its IMF components.
In our algorithm, the IMFs with peak frequency, $F_i$, in range of $[2, 12]$ Hz are categorized as a component of the tremor activity. Then, the filtered tremor signal can be reconstructed as [39]:

$$T_{ref}(t) = \sum_i h_i(t) \quad (F_i \in [2, 12] \text{ Hz}) \quad (4.7)$$

As depicted in Figure 4.10, the first three IMFs have a peak frequency in the range of $[2, 12]$ Hz. Therefore, the tremor activity of the signal can be extracted from only these three IMFs.

Figure 4.11 illustrates the reconstructed signal based on using IMFs with dominant frequency components in the range of $[2, 12]$ Hz (first three IMFs). As shown, the noise and voluntary motion of the hand is removed in the reconstructed signal.

Figure 4.10: PSD of the IMF components of the tremor recording.
4.7 Handedness of Tremors

Upon further investigation of the tremor signals, the relation between Alcohol Withdrawal Syndrome (AWS) tremors in the left and right hands of patients was evaluated. In this step, 122 recordings were analyzed, captured from 61 patients presented to the emergency department.

The variations in frequency between the left and right hands of the same AWS patient were studied. By frequency, we refer to the mean peak frequency in the time-frequency image of the recording, as described earlier in the chapter. Figure 4.12 illustrates a relatively high correlation between the left and right hand mean peak frequencies with a correlation coefficient of 0.63. In general, a higher left hand frequency corresponds to a higher right hand frequency, indicating that AWS tremors generally occur at similar frequencies in both hands for real patients.
Figure 4.12: Relation between left hand and right hand tremor frequencies.

It is interesting to note that while there is a general correlation between cross-hand tremor frequencies, this is not always the case. On average, a variation of 1.2 Hz between the left and right hand tremor frequencies was found.

Next, an expert panel of three senior physicians was used to provide a ground truth rating for each recording. Figure 4.13 illustrates the relationship between the expert tremor ratings for the left and right hands. As shown, there is a strong correlation between the left and right hand expert ratings with a correlation coefficient of 0.923. In all but one of the 61 cases, the ratings for different hands varied at most by 1 point. It is questionable whether there is a biased motivation for expert raters to rate different hands similarly, since physicians might know it is the same person.

Based on the frequency analysis explained earlier in this chapter, the CIWA-Ar tremor ratings for each recording was estimated and the relationship between the left and right hand rating was evaluated. Figure 4.14 illustrates this relationship, showing a generally high correlation between the ratings for each hand (with a correlation coefficient of 0.852).
Figure 4.13: Relation between left hand and right hand CIWA-Ar tremor ratings by an expert panel of three senior physicians. Please note that there are a total of 61 points in the above graph, with most points overlapping at the same location.

Figure 4.14: Relation between estimated left hand and right hand CIWA-Ar tremor ratings using the logarithmic model described in Section 4.5.
The effect of handedness on evaluation of the severity of the AW tremor as well as the combination method to combine the score given to each hand in order to come up with a final severity estimate is discussed in Chapter 5.

4.8 Summary

In this chapter the following topics were discussed in the Time-Frequency analysis of AW tremors:

- Applying STFT to measure the energy of the tremor signal in different frequency ranges.
- Visualizing tremor signals in time and frequency domains using spectrogram.
- Describing the structure of the application used to collect tremor signals from patients with AW in emergency departments and data collection method.
- Finding the mean peak frequency of the real tremors in order to distinguish between factitious and real AW patients.
- Quantifying the severity of AW tremors using a logarithmic model.
- Investigating the capability of the EMD algorithm in removing noise and voluntary movement in tremor recordings.
- Illustrating few examples on how this noise removal algorithm would work on tremor signals with different severities.
- Investigating the band-limited energy and frequency of the tremor signals in both hands of patients with different handedness and finding the correlation between AWS tremors in the left and right hand of patients.
Figure 4.15: Outline of tremor assessments used in different stages of the analysis.
† Two patients had assessments from both of their hands. In each case, one assessment was categorized as Real Characteristic Tremor (mild) and the other was categorized as Real Characteristic Tremor (moderate/severe).
Chapter 5

Performance Evaluation and Clinical Validation

The focus of this chapter will be a discussion of different statistical measures used in evaluating the performance of the proposed model in quantifying the severity of AW tremors. The analysis process from the data collection step to the final estimation of the AW tremor severity score is also described in this chapter. Furthermore, the variability and accuracy of the logarithmic model described in the previous chapter is explored by comparing it with the AWS tremor rating provided by junior and senior nurses, and physicians in the emergency department. An investigation into the best way to combine the information gathered from both hands is conducted to provide an accurate tremor severity estimate.

Additionally, the reliability of accelerometer chips on iOS devices is discussed for use in the medical applications as well as the variability of the accelerometer recordings in different iOS platforms. Lastly, the logarithmic model is validated in estimation of the severity of AW tremors in a clinical setting.

5.1 Statistical Analysis

A number of statistical measures have been employed in this thesis to measure the effectiveness of the proposed model in quantifying severity of alcohol withdrawal tremors on an 8-point scale.
5.1.1 P-Value

In discussing null hypothesis, the null hypothesis is to be rejected and therefore the hypothesis of equality is to be rejected. In analysis, the one-sided p-value was used with a significance level ($\alpha$) set to 0.05. Significance level is the probability of rejecting the null hypothesis given that it is true. It is usually set at or below 5%. The p-value is calculated from Mann-Whitney U test of medians.

5.1.2 Confidence Interval

Setting the significance level to 0.05, resulted in a exploration of 95% confidence interval in the analysis. Also in order to calculate the confidence interval, the Hodges-Lehmann’s method was used. A 95% confidence interval has a 95% probability of containing the population mean and it contains 95% of the population.

5.1.3 Root Mean Square Error

The Root Mean Square Error (RMSE) was used to measure the accuracy of the tremor severity estimation based on the proposed logarithmic model versus the ratings from the panel of expert physicians. The average RMSE was used to measure the inter-rater variability of tremor severity estimation within different groups of raters.

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (S_i - C_i)^2}$$

where $S_i$ is the estimated tremor severity rating for the $i^{th}$ patients and $C_i$ is the consensus rating for the $i^{th}$ patients from our panel of three expert physicians.

5.1.4 Cohen’s Kappa Agreement Measure

The inter-observer variation can be measured in cases that have different observers or raters evaluating the same experiment [52], [53]. In this thesis, there are different groups of raters, junior and senior physicians and nurses, our panel of three expert physicians, and the rating provided by our logarithmic model.

The Cohen’s Kappa agreement [54] calculates how much agreement is actually present compared to how much would be expected to be present by chance. Suppose that two methods $X$ and $Y$ are categorizing $n$ subjects into $m$ different categories. Let $f_{ij}$ denote the frequency of the number of subjects who put into category $i$ based on method $X$ and
method $Y$ categorized them as $j$. Table 5.1 illustrates the agreement table in our example. Please note that the agreement table would be different if the assumption about the agreement changes.

Table 5.1: Agreement table.

<table>
<thead>
<tr>
<th>$X = C_1$</th>
<th>$Y = C_1$</th>
<th>$Y = C_2$</th>
<th>$Y = C_3$</th>
<th>...</th>
<th>$Y = C_m$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X = C_1$</td>
<td>$f_{11}$</td>
<td>$f_{12}$</td>
<td>$f_{13}$</td>
<td>...</td>
<td>$f_{1m}$</td>
</tr>
<tr>
<td>$X = C_2$</td>
<td>$f_{21}$</td>
<td>$f_{22}$</td>
<td>$f_{23}$</td>
<td>...</td>
<td>$f_{2m}$</td>
</tr>
<tr>
<td>$X = C_3$</td>
<td>$f_{31}$</td>
<td>$f_{32}$</td>
<td>$f_{33}$</td>
<td>...</td>
<td>$f_{3m}$</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>$X = C_m$</td>
<td>$f_{m1}$</td>
<td>$f_{m2}$</td>
<td>$f_{m3}$</td>
<td>...</td>
<td>$f_{mm}$</td>
</tr>
</tbody>
</table>

The observed proportional agreement between $X$ and $Y$ is defined as:

$$p_o = \frac{1}{n} \sum_{i=1}^{m} f_{ii} \quad (5.2)$$

and the expected agreement by chance is:

$$p_e = \frac{1}{n^2} \sum_{i=1}^{m} f_{i+} f_{+i} \quad (5.3)$$

where $f_{i+}$ is the sum of the $i^{th}$ row and $f_{+i}$ is the sum of the $i^{th}$ column. The Kappa statistic is:

$$\hat{k} = \frac{p_o - p_e}{1 - p_e} \quad (5.4)$$

Kappa is standardized on a ‘−1’ to ‘1’ scale, where 1 is perfect agreement and ‘0’ is what would be expected by chance agreement.

5.2 Analysis Process

The process of estimating the severity of AW tremors using the logarithmic model and the flow of the application that is now being used by the emergency department are illustrated in Figure 5.1 and Figure 5.2, respectively.
5.2.1 Frequency Analysis

The frequency analysis in this thesis has three main steps:

1. Filtering the noise and voluntary movement from the recordings.

2. Finding the frequency range of AW tremors.

3. Building the model on clean tremor signals based on the band-limited energy of the AW tremors.

In order to remove the noise in the tremor recordings, different frequency ranges for the AW tremors were explored. Based on the literature of tremor analysis and also a preliminary analysis on the accuracy of the model by eliminating different frequency ranges, the peak frequency in the range of [2, 12] Hz was chosen to be our range of interest and anything outside this range was removed using Empirical Mode Decomposition algorithm described in Chapter 4.
After removing inherent noise in the signal, an investigation into the best frequency range to consider was conducted while estimating severity of AW tremors. Different frequency ranges \([f_0, 15]\) Hz were considered as well as measurements of the accuracy of the model obtained from the respective band-limited energy. 15 Hz was chosen as the upper limit for two main reasons:

1. From the AW tremor recordings in the data set, not much energy was observed above 15 Hz.

2. Based on the literature of tremor analysis, the peak frequency of the neurological tremors in most cases does not exceed 12 Hz.

Based on the analysis described in Chapter 4, \(f_0 = 5\) Hz came out to be the best lower limit for the frequency range.

### 5.3 Effect of EMD in Improving the Logarithmic Model

Using the EMD algorithm, one can extract the involuntary part of the tremor signal in order to see if the logarithmic model can be obtained to quantify the severity of the tremor more accurately.
In summary, the steps of finding the logarithmic model are:

1. Applying EMD to remove noise and voluntary movement from the original signal. Steps are shown in Figure 5.3.

2. Compute STFT with a Hanning window of 4-seconds and 90% overlap.

3. Compute the energy of the signal in each of the windows in range of $[5, 15]$ Hz. Energy of the original signal is the summation of the energy of all of the 40 windows.

4. Fit a logarithmic model of the form:
   \[ \log(1 + b.E) \]
   where ‘$E$’ is the energy of the signal calculated in step 2 and ‘$a$’ and ‘$b$’ are constants. This model is forced to pass through the origin when energy of the tremor is zero. Details of fitting the proper logarithmic model were described in Section 4.5.

### 5.3.1 Experimental Results

A 3-fold cross-validation method was used to measure the accuracy of the logarithmic model after applying the EMD algorithm to extract the tremor component of the signal. To accomplish this, the cross-validation was ran for 20 iterations on 117 recordings. The average RMSE of the tremor score computed by the algorithm compared with the consensus rating from a panel of three physicians was 0.71.

In order to determine the effectiveness of the extraction of the tremor component of the signal compared to using the original recording, the logarithmic model was trained on the original recording in 20 iterations of the 3-fold cross-validation. This time, the average RMSE of the tremor score computed by the algorithm, compared to the consensus rating, was 0.96.

### 5.3.2 Example of Removal of Voluntary Movements

One example of a tremor recording that involves voluntary movement is demonstrated in this section. In the spectrogram from the original tremor recording, there is a significant amount of noise involved in the recording (Figure 5.4), whereas in the clean signal, after the application of the EMD algorithm, it is evident that the main tremor activity in the signal appears around 8 Hz (Figure 5.5).
5.3.3 Evaluation of Performance of EMD depending on Tremor Severity

The experimental results show that depending on the severity of the tremor, the effect of the EMD algorithm in removing noise and voluntary movement would be different. Using the EMD algorithm as the noise removal mechanism in removing voluntary movement works best for tremors with a consensus rating between 2 and 5. The detailed performance of the EMD algorithm is shown in Table 5.2.

There could be different explanations as to why the performance decreased in the categories of 0 and 1 consensus severity rating and also the severity score of 6. One
Figure 5.4: Spectrogram of the original tremor recording from a recording involving voluntary motion.

explanation is that in estimating the severity of a tremor for these two tremor severity ranges by itself is very difficult, either you do not have much tremor component in the recording to rely on or the intensity of the developed tremor is too abundant in high frequency ranges. In the first instance, one might rely too heavily on the very small energy found in the range of [5, 15] Hz, and in the second case, one would have to remove part of the tremor component by excluding some IMFs and both result in a decrease in accuracy of the estimation.

In the following, you can see some cases of tremors from different severity scales and the performance of the EMD in removing voluntary movements. For each of the categories of ‘No Observable Tremor’, ‘Mild Tremor’, and ‘Severe Tremor’, are illustrated the original signal, what is considered noise and voluntary movement by the EMD algorithm, and the filtered signal after removing voluntary movement.
5.3.3.1 Example of ‘No Observable’ Tremor:

This example is recorded from a patient with no or limited observable tremor with a tremor severity score of 0. As shown in Figure 5.6, we have the original tremor recording with some frequency elements in higher frequency ranges around second 5 of the recording, as well as some low frequency elements for the whole duration of the recording.

Figure 5.7 shows the voluntary movement component, which will be removed as a result of applying the EMD algorithm. It mainly has low frequency components. Also, the filtered signal in Figure 5.8 shows no observable tremor in low frequency and a single spike of high frequency at around second 5 of the recording.
Table 5.2: The effect of EMD noise removal on severity estimation of different groups of tremors.

<table>
<thead>
<tr>
<th>Consensus Tremor Rating</th>
<th>Average RMSE before EMD</th>
<th>Average RMSE after EMD</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1.8</td>
<td>2.2</td>
</tr>
<tr>
<td>1</td>
<td>0.37</td>
<td>0.58</td>
</tr>
<tr>
<td>2</td>
<td>0.59</td>
<td>0.53</td>
</tr>
<tr>
<td>3</td>
<td>1.05</td>
<td>0.83</td>
</tr>
<tr>
<td>4</td>
<td>0.54</td>
<td>0.43</td>
</tr>
<tr>
<td>5</td>
<td>0.88</td>
<td>0.85</td>
</tr>
<tr>
<td>6</td>
<td>0.53</td>
<td>0.67</td>
</tr>
</tbody>
</table>

Figure 5.6: Original tremor recording.
Figure 5.7: Voluntary movement in tremor recording.

Figure 5.8: Filtered tremor by removing voluntary movements.
5.3.3.2 Example of ‘Mild’ Tremor:

This is an example of a mild tremor, which was rated as a 3 out of a maximum of 7. As shown in Figure 5.9, the original tremor recording has frequency components covering a range of [0, 15] Hz at the beginning of the recording. By finding the voluntary movement shown in Figure 5.10 you can see that most of the frequency components for the first few seconds are due to the voluntary movement of the patient.

As shown in Figure 5.11, the dominant frequency in a clean tremor signal is between [7, 10] Hz during the 20 seconds of the recording.

![Figure 5.9: Original tremor recording.](image-url)
Figure 5.10: Voluntary movement in tremor recording.

Figure 5.11: Filtered tremor by removing voluntary movements.
5.3.3.3 Example of ‘Severe’ Tremor:

Here is an example of a severe tremor, which received a tremor rating of 6. In the original recording shown in Figure 5.12 one can observe a dominant frequency around 8 Hz, but can also see some low frequency components at the beginning of the recording. Caught by the EMD algorithm, those low frequency components were the result of voluntary movement by the patient, potentially changing hand position (Figure 5.13).

The clean signal is shown in Figure 5.14 with one major dominant frequency around 8 Hz.
Figure 5.13: Voluntary movement in tremor recording.

Figure 5.14: Filtered tremor by removing voluntary movements.
5.4 Effect of EMD on Classification of Tremors

In Chapter 4, the classification of different types of tremors was explored in order to identify factitious tremors. This section will take a closer look at the possibility of classifying tremors after applying the EMD algorithm.

The EMD algorithm is applied to remove voluntary movement and noise captured in the tremor recordings. Then, the energy of the signal is investigated in the frequency range of [5, 15] Hz as well as the mean peak frequency, both described in Chapter 4. As plotted in Figure 5.15, the energy of the signal is shown in [5, 15] Hz vs. mean peak frequency for 3 types of tremors; ‘factitious tremor from patients’, ‘factitious tremor from nurses’, and ‘real tremors’.

![Figure 5.15: Energy in [5, 15] Hz vs. mean peak frequency of tremors for 3 types of 'factitious tremors', 'nurses' tremors', and 'real tremors (moderate/severe)'.](image)

As shown, most of the factitious tremors from patients have low energy and mean peak frequency less than 7 Hz. Additionally, looking at the factitious tremor from nurses, the total energy in the frequency range of [5, 15] Hz does not exceed 100. This confirms the argument made in Chapter 4 that mean peak frequency can be an informative feature in distinguishing factitious vs. real tremors. By removing the noise from the signal, we
found out that energy can also be another good feature.

Alternatively, if you look at the green points on the plot, they mostly have energy higher than 100 and mean peak frequency above 7 Hz.

Another important aspect of tremor classification is classifying real tremors into 3 groups of ‘No Observable Tremor’, ‘Mild Tremor’, and ‘Severe Tremor’ as the core classes physicians refer to. Figure 5.16 illustrates energy of the signal in [5, 15] Hz after applying the EMD to remove noise vs. mean peak frequency of the clean signal.

Based on Figure 5.16 in 82% of the cases the mean peak frequency of the signal is above 7 Hz. This observation further supports the idea of using a frequency of 7 Hz as a cut-off for finding factitious tremors. Notably, there is a clear distinction between the 3 types of tremors based on energy in [5, 15] Hz. Energy of less than 25 corresponds to the class of ‘no observable’ tremor, the energy in the range of 25 to 75 most likely refers to ‘mild’ tremor, and energy above 75 corresponds to ‘severe’ tremor. It should be noted that this classification was given by the panel of physicians by looking at the video recordings from patients during assessment. The illustration in Figure 5.16 supports the original idea of using the energy of the signal to quantify severity of real tremors and in fact after application of the EMD algorithm the boundary between different classes is more evident.

5.5 Analysis of the Logarithmic Model

In Chapter 4, a logarithmic model was proposed based on the energy of the tremor signal in the frequency range of [5, 15] Hz. The training step includes 84 recordings from 50 patients. This section looks into testing the model on another set of 33 recordings from 33 patients.

5.5.1 Testing the Model

The logarithmic model found in the training step was tested on a separate set of 33 patients (33 recordings). In this set, 22 recordings were categorized as ‘real, characteristic, mild’ tremors, 7 were categorized as ‘real, characteristic moderate/severe’ tremors, and 4 were categorized as ‘no observable’ tremors. The proposed logarithmic relationship provides a CIWA-Ar Root Mean Square Error of 0.91 compared to the consensus rating from expert physicians on the test set.
5.5.2 Error of the Proposed Logarithmic Model

Adding all the tremor signals in the training and test sets together, the frequency range with the minimum RMSE remains at [5,15] Hz. Figures 5.17 and 5.18 illustrate the RMSEs for different frequency ranges and the logarithmic model based on the combined total of 117 recordings from 83 patients. The logarithmic model trained on the the training and test sets using non-linear regression algorithm is of the form $S = 4.99 \log(1 + 0.062E)$ with $S$ and $E$ being the estimated tremor severity rating and the band-limited energy in the range of [5, 15] Hz, respectively.

5.5.3 Inter-rater Variability

The average RMSE and kappa agreement within 1 point between each group of raters and consensus ratings assigned to tremors by expert physicians is calculated and presented in Table 5.3. This table demonstrates that the error in estimation of the tremor severity, based on the logarithmic model, is inline with those of senior physicians and
Chapter 5. Performance Evaluation and Clinical Validation

Figure 5.17: RMSEs for different frequency ranges based on the training and test sets.

Comparing the kappa agreement within 1 point between all raters and the consensus rating, it appears that the logarithmic model has a higher agreement with the consensus rating even without applying EMD noise removal.
Figure 5.18: Relation between consensus CIWA-Ar tremor score and energy of the tremor signal in [5, 15] Hz ($S = 4.99 \log(1 + 0.062E)$) based on the training and test sets.

Additionally, the average RMSE between each group of raters is investigated along with consensus ratings for different tremor rating ranges: low-range tremor ratings (0 to 2), mid-range ratings (3 to 4), and high-range ratings (5 to 7). Table 5.4 presents different RMSEs for different ranges of tremor severity.

From Tables 5.3 and 5.4 it can be concluded that the logarithmic model based on original tremor recordings performs better than all groups of raters in mid and high range ratings and the accuracy of the model is inline with physicians in low range ratings.

A comparison was made between the model’s ability to rate the tremor severity on the CIWA-Ar scale and the ability of nurses and physicians with different levels of experience, with consideration that the ratings provided by clinicians were obtained in an artificial setting. Clinicians were asked to rate all 33 videos sequentially, and were therefore able
Table 5.3: Inter-rater variability.

<table>
<thead>
<tr>
<th>Group of raters</th>
<th>Average RMSE compared to the consensus rating</th>
<th>Kappa of agreement within 1 point</th>
</tr>
</thead>
<tbody>
<tr>
<td>Junior Nurses</td>
<td>0.94</td>
<td>0.61</td>
</tr>
<tr>
<td>Senior Nurses</td>
<td>0.94</td>
<td>0.58</td>
</tr>
<tr>
<td>Junior Physicians</td>
<td>1.04</td>
<td>0.4</td>
</tr>
<tr>
<td>Senior Physicians</td>
<td>0.95</td>
<td>0.48</td>
</tr>
<tr>
<td>Estimation Based on</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proposed Logarithmic Model</td>
<td>0.91</td>
<td>0.76</td>
</tr>
<tr>
<td>Proposed Logarithmic Model</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model without EMD</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model with EMD</td>
<td>0.79</td>
<td>0.82</td>
</tr>
</tbody>
</table>

Table 5.4: Average RMSE for different groups of raters in different tremor rating ranges.

<table>
<thead>
<tr>
<th></th>
<th>Low-range</th>
<th>Mid-range</th>
<th>High-range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Junior Nurses</td>
<td>0.90</td>
<td>1.03</td>
<td>0.94</td>
</tr>
<tr>
<td>Senior Nurses</td>
<td>0.91</td>
<td>0.96</td>
<td>1.03</td>
</tr>
<tr>
<td>Junior Physicians</td>
<td>1.00</td>
<td>1.15</td>
<td>1.01</td>
</tr>
<tr>
<td>Senior Physicians</td>
<td>0.98</td>
<td>0.98</td>
<td>0.91</td>
</tr>
<tr>
<td>Estimation Based on</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proposed Logarithmic Model</td>
<td>0.95</td>
<td>0.83</td>
<td>0.73</td>
</tr>
</tbody>
</table>

Standardization of the tremor assessment is the key element of the CIWA-Ar application, as it is the clinical symptom most commonly used to quantify withdrawal severity, and subject to widely variable interpretation based on clinical experience. Two random-
ized controlled trials have shown that when AWS patients are treated using a CIWA-Ar based symptom-guided approach their treatment is improved (faster symptom resolution, lower total administered doses of benzodiazepines) compared to other approaches [9], [17]. In an environment in which demand for ED stretchers regularly exceeds availability, any strategy that both improves care and efficiency deserves further exploration.

5.6 Combining Severity Scores for Both Hands

In Chapter 4, the effect of handedness was considered in relation to the mean peak frequency and the severity of tremors. Another very important and practical question involves the method by which data from different hands should be combined to obtain an overall rating estimate. For example, if in the emergency room two recordings are made from both hands, which recording should be used and if both are used, should they be combined? To answer this, the Root Mean Square Error was measured for five different cases, with the results shown in Table 5.5.

The RMSE is a measure of how much the tremor rating estimate varies from the expert rating. As compared to an absolute error measure, RMSE places a higher weight on larger errors and provides a less conservative error estimate.

Table 5.5: Left-Right tremor data combination methods and their accuracy in estimating CIWA-Ar tremor score.

<table>
<thead>
<tr>
<th>Data Combination Method</th>
<th>CIWA-Ar Tremor RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Left-Right Max</td>
<td>0.985</td>
</tr>
<tr>
<td>Left-Right Min</td>
<td>1.06</td>
</tr>
<tr>
<td>Right Only</td>
<td>0.984</td>
</tr>
<tr>
<td>Left Only</td>
<td>1.111</td>
</tr>
<tr>
<td>Left-Right Average</td>
<td>0.977</td>
</tr>
</tbody>
</table>

The Left-Right Max method separately estimates the tremor score for each hand and the maximum is used to compare to the maximum rating of the expert panel. This is essentially using the worst tremor to define the tremor rating, and leads to a generally lower RMSE of 0.985. The Left-Right Min method uses the lowest tremor rating as the
estimate, and generally has a higher RMSE.

It is interesting to note the difference between the Right Only and the Left Only tremor rating methods, with the Right Only resulting in a considerably lower RMSE than the Left Only method. Could this be because most of the participants were right handed and were able to produce a more accurate tremor score from their dominant hand? This is possible, but is not explored further.

The lowest RMSE was obtained by averaging the estimated tremor ratings for the left and right hands (and comparing this to the expert panel average rating for the two hands). The motivation here is that a slight overestimation or underestimation on either hand can be overcome by using more data from the other hand to obtain an even more accurate tremor assessment. There could, however, be another explanation for this lower RMSE. It could be that the ground truth tremor rating (obtained from the expert panel) is more accurate when averaged across both hands than when taken from either hand alone.

Figure 5.19 illustrates the relationship between the left-right tremor rating averages as obtained from the logarithmic model explained in Section 4.5 and from the expert panel. As shown, there is a strong correspondence between the estimated tremor rating and the expert tremor rating when both are averaged for both hands. This suggests that taking recordings from both the left and right hands and averaging the estimated ratings is a practical way of obtaining a single overall CIWA-Ar tremor rating for clinical use.

At this point, it is useful to reflect on the importance of the tremor frequency on the overall ratings. One simple evaluation is to look at the average frequency for the highest-rated hand versus the average frequency for the lowest-rated hand. For the highest-rated hand, the average frequency was 6.55 Hz, while for the lowest-rated hand the average frequency was 6.1 Hz. The average frequency for the left hand was 6.26 Hz and for the right hand it was 6.39 Hz. Consequently, there seems to be a weak, if any, relationship between higher frequencies and higher-rated tremors.

Alternatively, the impact of frequency on the tremor rating RMSE is rather interesting. Figure 5.20 evaluates this relationship, showing how the estimated tremor rating RMSE varies for tremors in different frequency ranges.

As shown, for both the Left-Right Max and Left-Right Average methods, the lowest RMSE is obtained at the highest frequency range. This demonstrates that the most accurate tremor estimates are obtained at higher frequencies. For the Left-Right Average method, there is a decreasing relationship between RMSE and frequency with a high
Figure 5.19: Relation between left-right averaged estimated tremor rating and the left-right averaged expert tremor rating.

RMSE of 1.102 for the 3-6Hz range and a low RMSE of 0.625 for the 7-10Hz range.

An analysis of these results is required. First, the tremor rating improvements obtained by averaging the two hands was only marginally better. This is especially true since the scores for the two hands are usually very similar. Second, there is a potential bias in the physician ratings of each hand since they can see both hands during the rating process, which could bias the results towards an average rather than a maximum or a minimum.

5.7 Accelerometer Signal Variability in Different Devices

It is essential to note that there is a wide variation across different devices. Accelerometer chips vary across device types. For example, 4th and 5th generation iPod Touch devices are reported to use the “STMicroelectronics LIS331DLH” accelerometer, which measures acceleration in units of g, and gives a maximum frequency of 100
Figure 5.20: Relation between estimate tremor rating and tremor frequency for (black) the Left-Right Max method and (grey) the Left-Right Average method.

Hz [61]. In comparison, iPhones 5 and 5s are reported to use the “Bosch Sensortech BMA220” 3-axis accelerometer [85] and iPhones 6 and 6+ have “InvenSense MP67B” 6-axis Gyroscope and Accelerometer Combo and a “Bosch BMA280” 3-axis accelerometer. These different accelerometer chips have different performances and accuracies [86], [87].

The mobile application was tested on the iPod touch 5th generation, as well as the 5s, 6, and 6+ iPhone models using the platform generating consistent acceleration. The readings obtained from different iOS devices are generally different for the same acceleration. This is partially due to the size and geometry of the devices, but likely also dependent on the different accelerometer chips used in each device.

To understand these issues, two experiments were set up. First, a weak-vibration experiment was set consisting of a pendulum-connected platform that would vibrate at approximately 2 Hz. The following is the average of 20 trials using the logarithmic tremor severity scale described in Chapter 4. Tables 5.6 and 5.7 summarize the average and standard deviation of the ratings obtained from different iOS platform.
Table 5.6: Average rating and rating standard deviation in different iOS devices for the weak vibration platform.

<table>
<thead>
<tr>
<th>Device</th>
<th>Average Rating (0 - 7 scale)</th>
<th>Rating Std. Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>iPod (5th Generation)</td>
<td>1.38</td>
<td>0.09</td>
</tr>
<tr>
<td>iPhone 5s</td>
<td>1.01</td>
<td>0.10</td>
</tr>
<tr>
<td>iPhone 6</td>
<td>0.87</td>
<td>0.12</td>
</tr>
<tr>
<td>iPhone 6+</td>
<td>0.89</td>
<td>0.17</td>
</tr>
</tbody>
</table>

Keeping in mind the variation in devices, the results are fairly consistent when using the same device, which can be seen by the low standard deviation. Also, the iPhone 6 and 6+ have the same chipset but different sizes. Despite the size difference, the two devices have a very similar average reading, indicating that the variations above are more likely due to the electronics than the mechanics of the device.

To assess the performance with a stronger vibration, a second experiment was set up with an intense platform that was directly moving up and down at approximately 5 Hz. Figures 5.21 and 5.22 demonstrate the first and second platforms, respectively. The following Table 5.7 shows the results based on 20 trials of the same logarithmic model on different devices using the platform with high intensity vibrations.

Table 5.7: Average rating and rating standard deviation in different iOS devices for the strong vibration platform.

<table>
<thead>
<tr>
<th>Device</th>
<th>Average Rating (0 - 7 scale)</th>
<th>Rating Std. Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>iPod (5th Generation)</td>
<td>6.53</td>
<td>0.34</td>
</tr>
<tr>
<td>iPhone 5s</td>
<td>5.08</td>
<td>0.36</td>
</tr>
<tr>
<td>iPhone 6</td>
<td>6.22</td>
<td>0.33</td>
</tr>
<tr>
<td>iPhone 6+</td>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>

It must be noted that no data was recorded using the iPhone 6+ on the strong-intensity vibration platform because the weight and size was above the handling limit of our platform. The standard deviation is fairly consistent for the different devices indicating that substantial variations are most likely due to the electronics and/or mechanics of the device rather than experimental error. Finally, whereas the iPhone 5s and the iPod
(5th Generation) had similar results in the weak-vibration experiment, in this second strong-vibration experiment their results vastly varied.

5.7.1 Model Correction

Knowing that accelerometer readings vary from device to device depending on the accelerometer chip and the physical properties of the device, the next goal is to modify the model to adapt to different device types. Based on the readings collected from the iPod 5th generation, iPhone 5s, and iPhone 6 shown in Tables 5.6 and 5.7, a preliminary model correction was made for the iPhone 5s and iPhone 6.

In Figure 5.23 two data points were used per device type, one corresponds to the average rating from a low frequency platform and the other corresponds to the average rating from a high frequency platform, both compared to the rating from iPod 5th generation, which was the main device used for the data collection in the emergency department. The corrected model for the iPhone 5s based on the iPod 5th generation is a linear model of the form:

\[ R_{5s} = 0.77 \times R_{iPod} \]  \hspace{1cm} (5.5)
and the corrected model for iPhone 6 is:

\[ R_6 = 0.6 \times (R_{iPod})^{1.25} \]  \hspace{1cm} (5.6)

It must be noted that the corrected models for the iPhone 5s and iPhone 6 have not been clinically verified like the iPod 5th generation. These models are just based on a few data points collected to verify variability of the accelerometer readings across different device types.

With the above, the difference in tremor readings varied by at most 1.45 points for an intense vibration and 0.51 points for a weak variation. Therefore, even if recordings are being collected with different device types, the classification of the tremor is still within 0.5 - 1.5 points of that of the experts, which is often better than clinical judgment. However, a better solution is to study the rating performance of each device and adjust the scores to provide a universal score with all device biases removed. There will remain errors, and this will not be as reliable as a single-device based assessment, but at the
very least, it will limit the cross-device assessment errors.

In conclusion, while using one device type for data collection and model validation the experimental error is negligible. Additionally, although it is possible to have a universal model with device biases removed, the focus was on validating the logarithmic model based on the data collected from an iPod 5th generation.

5.8 Clinical Validation of the Logarithmic Model

In this section of the research, the model will be validated in a clinical setting and measured for its accuracy compared to the expectation of the panel of three expert physicians. To achieve this, the model is being validated for a 10-month period in one of the emergency departments. Outlined here is the design of the clinical validation step of the thesis and its outcomes.

5.8.1 Study Design and Setting

The data collection for this research took place in the emergency departments of two large academic tertiary care hospitals in Toronto. Data for the derivation of the tremor model was obtained over a 15-month period (July 2013 - October 2014), and data for prospective validation of the model component was obtained from a single site over a 10-month period (October 2014 - August 2015). Both sites have annual patient census’s exceeding 60,000 visits each year, are geographically located in the downtown core of the Greater Toronto Area (population 6 million), and were staffed by clinicians with
postgraduate certification in emergency medicine. Both sites have extensive experience managing patients with alcohol withdrawal, including the use of treatment algorithms incorporating the CIWA-Ar for at least 5 years prior to the initiation of this project. This research was approved by the Institutional Review Board (IRB) at each hospital, and patients provided written informed consent at the time of enrolment.

5.8.2 Selection of Participants

Data was collected from a convenient sample of patients presenting to the ED of either hospital when research staff was available to enrol patients. To be eligible, participants had to be identified as clinically having alcohol withdrawal, have been placed on a CIWA-Ar protocol in the ED for evaluation and treatment, and be competent to provide consent for participation in English.

5.8.3 Methods and Measurements

A smartphone app version of the CIWA-Ar was developed and pilot tested for usability by clinical staff. Evaluation of the layout and interface design was discussed at a focus group meeting, which included a selection of interested end-users (MD’s and RN’s), and the feedback obtained was used to improve usability and clarity of the interface.

The app included an electronic tremor assessment using the built-in accelerometer of handheld iOS devices. The device quantifies and characterizes tremor severity by measuring the magnitude of accelerations in three dimensions. Seated patients were asked to place the device in the upturned palm of each hand (one at a time), with their arms fully outstretched and unsupported. Data was collected for over 20 seconds with a sampling rate of 65 Hz. Tremor was assessed on both hands, and each assessment was videotaped by the research assistants for later review by the panel. Each tremor assessment video was subsequently independently reviewed and rated (using the 0 - 7 CIWA-Ar scale) by three expert assessors. Individual assessments were compared across assessors, and any disparities were resolved by consensus using agreed upon reference examples to guide decision making.

5.8.4 Clinical Validation of the Logarithmic Model

In order to validate the ability of the logarithmic model in estimating the severity of AW tremors, research staff collected evaluations (obtained for use in the clinical care of
patients) of tremor data, using three methods:

1) Electronically, using the logarithmic model proposed in Chapter 4 in realtime.
2) The RN’s actual subjective tremor rating (blinded to the electronic assessment).
3) Video recordings of these assessments, which were subsequently independently viewed and rated by a panel of three expert clinicians, which resolved any discrepancies by consensus.

Expert consensus ratings were considered the gold-standard, and were compared to both the rating generated by the logarithmic model and the nurse’s subjective tremor rating.

5.8.5 Outcomes

The primary outcome of this research was the agreement of tremor severity assessments (using the 0 - 7 CIWA-Ar scale) between the logarithmic model and expert clinician consensus rating.

The secondary outcome was the agreement of tremor severity assessments (using the 0-7 CIWA-Ar scale) between the consensus rating from the panel of expert physicians, and clinical RN assessments performed as part of clinical care.

These two outcomes demonstrate that even though the consensus rating is based on the video recordings from the patients, rather than visiting them in-person, building a model based on the consensus rating will result in an agreement of the score provided by the model and the score provided by nurses visiting the patient at the same time.

5.8.6 Results

5.8.6.1 Demographics

Patient demographics for both the derivation and the internal validation portions of the research are included in Table 5.8. For the derivation and validation phases tremor data was collected from 117 individual assessments from 83 patients, and for the clinical validation of the model, an additional 78 tremor recordings were collected from 76 patients. For the comparison between the logarithmic model and the clinical RN rating, 62 tremor recordings from 31 individual patients were evaluated.
Table 5.8: Characteristics and emergency department demographics of included patients.

<table>
<thead>
<tr>
<th>Variable</th>
<th>N = 83</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>71 (85.5%)</td>
</tr>
<tr>
<td>Mean (SD) age (years)*</td>
<td>44.4 (11.6)</td>
</tr>
<tr>
<td>No fixed address*</td>
<td>11 (15.5%)</td>
</tr>
<tr>
<td><strong>Mode of Arrival</strong>*</td>
<td></td>
</tr>
<tr>
<td>Walk in</td>
<td>32 (45.1%)</td>
</tr>
<tr>
<td>Ambulance</td>
<td>39 (54.9%)</td>
</tr>
<tr>
<td><strong>CTAS score</strong>*</td>
<td></td>
</tr>
<tr>
<td>CTAS 2</td>
<td>31 (43.7%)</td>
</tr>
<tr>
<td>CTAS 3</td>
<td>40 (56.3%)</td>
</tr>
<tr>
<td><strong>Blood ETOH</strong>*</td>
<td></td>
</tr>
<tr>
<td>≤ 2</td>
<td>26 (36.6%)</td>
</tr>
<tr>
<td>Median (IQR) for ETOH &gt; 2</td>
<td>30.5 (17.8, 48.8)</td>
</tr>
<tr>
<td><strong>Previous history of ETOH-related seizures</strong>*</td>
<td></td>
</tr>
<tr>
<td>Median (IQR) number of Diazepam doses given in ED</td>
<td>3.0 (2.8, 3.2)</td>
</tr>
<tr>
<td>Median (IQR) total Diazepam given in ED</td>
<td>60.0 (27.5, 60.0)</td>
</tr>
<tr>
<td>Median (IQR) Diazepam sent with patient upon discharge</td>
<td>0 (0, 12.5)</td>
</tr>
<tr>
<td><strong>No history of ETOH-related seizures</strong>*</td>
<td></td>
</tr>
<tr>
<td>Median (IQR) number of Diazepam doses given in ED</td>
<td>3.0 (1, 4)</td>
</tr>
<tr>
<td>Median (IQR) total Diazepam given in ED</td>
<td>40.0 (7.5, 60.0)</td>
</tr>
<tr>
<td>Median (IQR) Diazepam sent with patient upon discharge</td>
<td>0 (0, 40)</td>
</tr>
<tr>
<td>Median (IQR) number of CIWA-Ar assessments in ED</td>
<td>4.0 (2.0, 6.0)</td>
</tr>
<tr>
<td>Median (IQR) ED length of stay (minutes)</td>
<td>385 (287, 575)</td>
</tr>
<tr>
<td>Admitted/ transferred</td>
<td>15 (18.1%)</td>
</tr>
</tbody>
</table>

Please note that in the Table 5.8, we are missing the demographics information for 12 patients because they are from the second site and the REB at in that site did not allow for us to review the charts of the patients enrolled. Therefore, all categories in the chart with a * are based off of the N value of 71 (without the 12 patients from the second site).

5.8.6.2 Derivation and Validation of the Logarithmic Model

Using the RMSE, the best frequency range fit (0.85) in which to collect accelerometer energy data was [5,15] Hz. The derivation model fit was found to be logarithmic and
computed according to the following relationship: \( R = 4.96 \log(1 + 0.067E) \), where \( R \) is the estimated CIWA-Ar tremor rating, \( E \) is energy of the accelerometer signal, and the log is base 10.

The agreement between the scores based on the logarithmic model and expert consensus ratings at each tremor intensity (0 - 7, rounded to the nearest integer) is shown in Table 5.9. There was a good agreement (kappa 0.89, 95% CI: 0.83, 0.96) between the logarithmic model and gold standard clinical experts at all tremor severities, with modal agreement within 1 point between the two agreeing in 107 out of 117 (91.4%) assessments.

Table 5.9: The logarithmic scores compared to the consensus expert tremor scores during the derivation of model.

<table>
<thead>
<tr>
<th>Tremor Score</th>
<th>n(%)</th>
<th>Logarithmic model Score</th>
<th>Agreement within 1 point</th>
<th>Agreement within &gt;1 point</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>14 (12.0%)</td>
<td>0 - 0.4</td>
<td>10</td>
<td>4</td>
</tr>
<tr>
<td>1</td>
<td>20 (17.1%)</td>
<td>0.5 - 1.4</td>
<td>20</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>30 (25.6%)</td>
<td>1.5 - 2.4</td>
<td>29</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>31 (26.5%)</td>
<td>2.5 - 3.4</td>
<td>27</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>12 (10.2%)</td>
<td>3.5 - 4.4</td>
<td>11</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>5 (4.3%)</td>
<td>4.5 - 5.4</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>5 (4.3%)</td>
<td>5.5 - 6.4</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>0 (0.0%)</td>
<td>&gt; 6.4</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Total</td>
<td>117</td>
<td>The overall kappa was very good 0.89 (95% CI: 0.83, 0.96)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Details of kappa agreement measure and confidence interval are described in Section 5.1.

5.8.6.3 Clinical Validation of the Tremor Model

The model deployed for the clinical validation step was based on the logarithmic model trained on the training and test sets explained in Section 5.5 without applying the EMD to remove the noise and voluntary movement in the tremor signal.

Of the 78 tremor recordings from 76 patients collected for clinical validation, logarithmic scale derived tremor scores matched exactly with expert assessor scores in 36 (46.2%) cases, within 1 point in 73 (93.6%) cases, and by 2 or more in 5(6.4%) cases. A comparison of model vs. expert consensus tremor scores at each tremor severity is shown
in Table 5.10. Overall, the agreement between the logarithmic model and the consensus expert scores being within 1 point was excellent, with a kappa of 0.92 (95% CI: 0.86, 0.99).

Table 5.10: The logarithmic scores compared to the consensus expert tremor scores during the validation of the model.

<table>
<thead>
<tr>
<th>Tremor Score</th>
<th>n(%)</th>
<th>Logarithmic model Score</th>
<th>Agreement within 1 point</th>
<th>Agreement within &gt;1 point</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>6 (7.7%)</td>
<td>0 - 0.4</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>15 (19.2%)</td>
<td>0.5 - 1.4</td>
<td>15</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>15 (19.2%)</td>
<td>1.5 - 2.4</td>
<td>14</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>22 (28.2%)</td>
<td>2.5 - 3.4</td>
<td>21</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>10 (12.8%)</td>
<td>3.5 - 4.4</td>
<td>9</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>5 (6.4%)</td>
<td>4.5 - 5.4</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>2 (2.6%)</td>
<td>5.5 - 6.4</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>3 (3.8%)</td>
<td>&gt; 6.4</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>78</td>
<td>The overall kappa was very good 0.92 (95% CI: 0.86, 0.99)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

5.8.6.4 Comparison of Tremor Score based on the Consensus Severity Rating and Clinical Nursing Assessments

Using 62 tremor recordings from 31 patients, RN-derived tremor scores matched exactly with expert assessor scores in only 11 (17.7%) instances, were within 1 point in 29 (46.8%), and differed by 2 or more points in 33 (53.3%) of cases. The kappa for agreement within 1 point for tremor severity was “fair” at 0.39 (95% CI: 0.25, 0.53). Details of the experiment are shown in Table 5.11.

Table 5.11: RN scores compared to the consensus expert tremor scores.

<table>
<thead>
<tr>
<th>Rater</th>
<th>n</th>
<th>Exact Agreement</th>
<th>Agreement ±1 point</th>
<th>Differ by 2+ points</th>
</tr>
</thead>
<tbody>
<tr>
<td>RN</td>
<td>62</td>
<td>11 (17.7%)</td>
<td>18 (29.0%)</td>
<td>33 (53.3%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>The overall kappa for agreement within 1 point was fair 0.39 (95% CI: 0.25, 0.53)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
5.8.7 Importance of Appropriate Tremor Severity Estimation

Retrospective medical records of adult patients presenting to an academic ED in alcohol withdrawal from July 2013 to January 2015 have been reviewed. Patient demographics, ED length of stay, administration of CIWA-Ar protocol (including the estimation of tremor severity), total dose of benzodiazepines administered in the ED, number of prescriptions and unit benzodiazepine doses given upon discharge were recorded.

Management of alcohol withdrawal was compliant with CIWA-Ar protocol in 38 of the 71 patients (53.5%). The common errors were overestimation of tremor severity and administration of benzodiazepine when it was not required.

Also, median (IQR) ED length of stay for patients whose management was compliant with CIWA-Ar was 5.8 (4.2 - 7.2) hours, while the median (IQR) length of stay for those patients with overestimation of tremor severity was 8.4 (5.2 - 11.4) hours.

Those patients with no protocol errors received lower total doses of benzodiazepine while in the ED, with no adverse outcomes. Table 5.12 summarizes the statistics and differences between the patients with appropriate CIWA-Ar protocol and those with overestimation of tremor severity.

Table 5.13 summarizes the demographics of patients from July 2013 and January 2015.
Table 5.12: Comparison of the ED length of stay and diazepam administration for cases of appropriate vs. inappropriate CIWA-Ar protocol.

<table>
<thead>
<tr>
<th></th>
<th>CIWA-Ar</th>
<th>Median</th>
<th>IQR</th>
<th>95% CI</th>
<th>p – value</th>
</tr>
</thead>
<tbody>
<tr>
<td>ED length of stay</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Appropriate</td>
<td>350 min</td>
<td>(253, 433)</td>
<td>153</td>
<td>(28, 253)</td>
<td>0.01</td>
</tr>
<tr>
<td>Inappropriate</td>
<td>503 min</td>
<td>(314, 681)</td>
<td>0</td>
<td>(-1, 1)</td>
<td>0.01</td>
</tr>
<tr>
<td>Number of Diazepam doses given in ED</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Appropriate</td>
<td>3 doses</td>
<td>(1, 3)</td>
<td>0</td>
<td></td>
<td>0.05</td>
</tr>
<tr>
<td>Inappropriate</td>
<td>3 doses</td>
<td>(3, 4)</td>
<td>(-1, 1)</td>
<td></td>
<td>0.05</td>
</tr>
<tr>
<td>Total Diazepam given in the ED</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Appropriate</td>
<td>40 mg</td>
<td>(7.5, 60)</td>
<td>20</td>
<td>(0, 30)</td>
<td>0.04</td>
</tr>
<tr>
<td>Inappropriate</td>
<td>60 mg</td>
<td>(30, 65)</td>
<td>(0, 30)</td>
<td></td>
<td>0.04</td>
</tr>
<tr>
<td>Diazepam sent with patient upon discharge</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Appropriate</td>
<td>0 mg</td>
<td>(0, 40)</td>
<td>10</td>
<td></td>
<td>0.04</td>
</tr>
<tr>
<td>Inappropriate</td>
<td>10 mg</td>
<td>(0, 40)</td>
<td>(0, 20)</td>
<td></td>
<td>0.04</td>
</tr>
<tr>
<td>Total Diazepam</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Appropriate</td>
<td>55 mg</td>
<td>(20, 73)</td>
<td>15</td>
<td></td>
<td>0.01</td>
</tr>
<tr>
<td>Inappropriate</td>
<td>70 mg</td>
<td>(60, 100)</td>
<td>(10, 40)</td>
<td></td>
<td>0.01</td>
</tr>
</tbody>
</table>

Table 5.13: Characteristics and emergency department demographics of included patients.

<table>
<thead>
<tr>
<th>Variable</th>
<th>N = 71</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>61 (85.9%)</td>
</tr>
<tr>
<td>Mean (SD) age (years)</td>
<td>44.1 (11.6)</td>
</tr>
<tr>
<td>No fixed address</td>
<td>11 (19.7%)</td>
</tr>
<tr>
<td><strong>Mode of Arrival</strong></td>
<td></td>
</tr>
<tr>
<td>Walk in</td>
<td>36 (50.7%)</td>
</tr>
<tr>
<td>Ambulance</td>
<td>35 (49.3%)</td>
</tr>
<tr>
<td><strong>CTAS score</strong></td>
<td></td>
</tr>
<tr>
<td>CTAS 2</td>
<td>31 (43.7%)</td>
</tr>
<tr>
<td>CTAS 3</td>
<td>39 (54.9%)</td>
</tr>
<tr>
<td>CTAS 4</td>
<td>1 (1.4%)</td>
</tr>
<tr>
<td><strong>Blood ETOH</strong></td>
<td></td>
</tr>
<tr>
<td>≤ 2</td>
<td>19 (26.8%)</td>
</tr>
<tr>
<td>Median (IQR) for ETOH &gt; 2</td>
<td>33.0 (17.5 , 56.5)</td>
</tr>
<tr>
<td><strong>Previous history of ETOH-related seizures</strong></td>
<td>28 (39.4%)</td>
</tr>
<tr>
<td><strong>CIWA-Ar Protocol Breaches</strong></td>
<td>33 (46.5%)</td>
</tr>
</tbody>
</table>
5.8.8 Discussion

The preceding section outlined the development of a model that reliably and objectively quantifies the severity of alcohol withdrawal related tremors, performing comparably to three clinical experts. Tremor being the most reliable and reproducible clinical sign of withdrawal, this model represents a potentially important improvement in the ability to accurately assess alcohol withdrawal severity, and care for patients with this condition.

Two randomized controlled trials have demonstrated that patients in alcohol withdrawal treated with a CIWA-based symptom guided protocol have faster symptom resolution, receive lower total doses of benzodiazepines, and have no increase in adverse outcomes (seizures, delirium) when compared to either ad-hoc or fixed dose regimens. On the strength of these trials, even though it was designed for use in the in-patient addiction treatment setting, the CIWA-Ar has become the gold-standard tool for assessing withdrawal severity in most clinical settings. The development of an objective, standardized, and reliable way of assessing the tremor of alcohol withdrawal is an important step in improving the accuracy and usability of the CIWA-Ar in a variety of clinical settings, including the ED. When combined with educational support regarding the identification and evaluation of alcohol withdrawal syndrome, the CIWA-Ar, including the accelerometer-based tremor assessment tool, can perform well and improve clinical decision making and care in the ED.

While the primary goal of this work was to improve the reliability of the CIWA-Ar score, it is important to recognize the limitations of the CIWA-Ar, which has prevented it from achieving widespread use in the ED. Even in centers where it is used regularly, there is often uncertainty regarding how to apply and score it. Factors contributing to this uncertainty relate to both differences in patient population, and the ED practice setting.

To begin with, in the chaotic ED environment there is often question of which patients should be placed on a CIWA-Ar protocol, with some believing it is a tool for diagnosing alcohol withdrawal. The subjective, and difficult-to-quantify nature of many of the elements of the CIWA-Ar only adds to this confusion. This results in a sense that clinicians, especially those without experience, lack confidence and feel they are simply making up numbers.

Secondly, the CIWA-Ar is time consuming to administer, taking up to 5 minutes or longer per assessment and symptom-driven protocol’s, which use the CIWA-Ar, often
recommend hourly assessments. In the ED of this study, where typical ED LOS for these patients range from 6 to 9 hours, CIWA-Ar assessments would consume between 30 and 45 minutes per typical patient stay. When added to the time to sign out and administer each dose of medication required, the total time required to perform and manage each assessment is considerable. In a clinical setting such as the ED, where there are many competing time sensitive tasks, it can be difficult for clinical staff to keep up with the assessments.

Finally, the relationship between patients and physicians in the ED environment is fundamentally different to that in an in-patient addiction treatment facility setting, for which the CIWA-Ar was developed. Patients in the latter are seeking help for their addiction, are likely willing to participate in their recovery, and are invested in accurately reporting their symptoms. Their progress can be observed for extended periods of time with less pressure to discharge the patient, and fewer competing priorities for the staff providing care. Addiction treatment staff will have significant experience in evaluating patients in withdrawal, and be more comfortable evaluating subjective complaints in ways that incorporate their own judgment to inform their rating scores. By contrast, patients are often brought to the ED intoxicated, detained for medical reasons without access to alcohol, and subsequently develop withdrawal. Most will have no interest in seeking treatment for their addiction.

The best medications for treating withdrawal, benzodiazepines, are highly addictive, frequently abused and diverted to the black market, and often sought for secondary gain. Patients who have previously been treated for alcohol withdrawal quickly learn that if they come to the ED endorsing alcohol withdrawal symptoms, they will likely be treated with benzodiazepines again. They may even receive a prescription for an additional ongoing treatment, creating a climate of distrust between physicians and patients.

The development of an accelerometer-based method of assessing the tremor of alcohol withdrawal represents potentially significant progress in addressing many of the above challenges. At a minimum, the ability to accurately quantify tremor severity contributes to the overall reliability of the CIWA-Ar, making it a more reliable and useful tool for general use in the ED. It will also improve the confidence of assessors in their clinical decision making. It may also help reduce variability in treatment, and allow a more rational approach to clinical decision making and resource allocation when deciding which patients require hospital admission, and the level of care required.

The ability to reliably assess tremor severity may also facilitate the development of a new, less time consuming and less subjective assessment tool to replace the CIWA-Ar
in the ED. This will require additional research to determine which of the individual elements of the CIWA-Ar contribute the most to its discriminatory value and result in the best performance. There are at least 17 different scoring systems, with a combined 32 different domains, which have been used to assess withdrawal severity. All have struggled with quantifying objective signs of withdrawal severity.

5.8.9 Limitations

There were several noteworthy limitations to this research. Firstly, we performed relatively few assessments in the high end of the CIWA-Ar tremor scale (6 and 7). This may have been because research staff were not available when these patients presented to the ED, patients may have been partially treated prior to assessment, and because severe withdrawal is a less common presentation in our community. Additionally, patients with severe alcohol withdrawal were disproportionately excluded from the study due to their inability to provide informed consent.

As well, time permitting, this research can be evaluated in a crossover design, in which two approaches of treating AW patients (using the logarithmic model and the subjective rating from nurses) are consecutively administered in each patient recruited in the study. This can help in separating treatment effect from the period effect [88].

5.9 Summary

In this chapter we outlined the main outcomes of the thesis. These outcomes come in five main categories:

- Examining the performance of our tremor severity estimation model with and without applying the EMD noise removal algorithm using 20 iterations of 3-fold cross-validation. This included a closer look at the performance of EMD on tremor signals with different severity levels.

- Testing and validating the logarithmic model on the test set and comparing the accuracy of the model with the inter-rater variability for different groups of raters.

- Combination of the severity estimates for both hands of patients to create with a single severity score for each patient.

- Variability of accelerometer signal in different devices.
- Validation of the model in a clinical setting over a 10-month period. The performance of the model was compared against the severity score provided by the panel of expert physicians.

Figure 5.24: CONSORT diagram of patients in the derivation and internal validation phase.
Chapter 6

Conclusion

Alcohol and alcohol-related illnesses are among the most common health issues resulting in emergency department visits around the world [2]. The ability to provide better, more efficient care for patients in AW has global implications.

This thesis has used signal processing techniques to first understand features of AW tremors such as peak frequency and energy. It was found that a real tremor signal has a mean peak frequency above 7 Hz. This can potentially be used to distinguish factitious from real tremors and avoid drug-seeking. Second, a logarithmic model was introduced to quantify the severity of AWS tremors. This model has been validated on a separate set of tremor recordings and the performance is compared with the junior and senior nurses and physicians. The variability of the tremor ratings was investigated in terms of the RMSE between the ratings provided by each group of raters compared to the consensus rating from a panel of three expert physicians. As well, there was a look at the RMSE between the severity score provided by the logarithmic model and the consensus rating. The average RMSE between the proposed logarithmic model and the consensus rating is 0.91 on an 8-point scale based on analysis of 33 tremor recordings. This accuracy is inline with that of senior nurses (0.94) and senior physicians (0.95).

Furthermore, in encountering with patients with different handedness, the effect of handedness was studied on estimating the severity of the tremor in each hand. Based on the analysis, there was a strong relationship between the tremor score provided by the panel of physicians and estimated tremor severity based on the model. Moreover, there was a strong correlation between the tremor score based on the logarithmic model of both hands. Finally, different methods of combining the data from the two hands was evaluated in order to obtain a single tremor rating estimate, and found that simply averaging the tremor ratings of the two hands results in the lowest tremor estimate error.
Since the model proposed relies on the data collected from the accelerometer, the reliability and variability of accelerometer chips from different iOS platforms was investigated for medical use. Five different devices were tested on two different platforms (low and high intensity vibrations). It appeared that due to different accelerometer chips and differences in the size and geometry of the device, the recorded signals are not consistent across all devices and need to be standardized. However, when looking at only one device, the recorded signals are consistent. Therefore, for the purpose of this thesis, one device was chosen as the reference device to collect all research data, with a preliminary model correction applied to remove device biases.

The EMD algorithm was examined to distinguish between voluntary and involuntary movements captured in the tremor signal of patients in alcohol withdrawal. The EMD algorithm was used to decompose the tremor signal into its IMFs. Each IMF represents a different physical notion. IMFs with peak frequency in the range of [2, 12] Hz were used to reconstruct the tremor component of the accelerometer recording.

To measure the performance of the method in extracting the tremor activity, the logarithmic scoring system employing our new technique was compared with the consensus rating from three expert physicians on an 8-point scale. Based on 20 iterations of 3-fold cross-validation on all recordings, it was found that the proposed method achieved an average RMSE of 0.71 with respect to the consensus rating, a significant improvement.

As for the final step, an extensive clinical validation study was conducted. The ability of the model to quantify the severity of AW tremors compared to the consensus rating was looked at. The agreement within 1 point between the model and consensus expert scores was excellent with a kappa of 0.92. Also, the analysis of the impact of providing an accurate tremor severity estimate on 71 patients presented to the ED from July 2013 to January 2015 shows that the median length of stay in the ED would have been decreased to 5.8 hours from 8.4 hours if the administration of CIWA-Ar protocol was appropriate. This improvement can potentially have a significant effect in improving the efficiency of running the ED.
Bibliography


Biomedical and Health Informatics (BHI), 2014 IEEE-EMBS International Conference on (pp. 113-116). IEEE.


Appendices
Appendix A

Medical Protocols

*Sweating

*Hallucinations*: “Are you feeling, seeing, or hearing anything that is disturbing to you? Are you seeing or hearing things you know are not there?”

*Orientation*: “What is the date, month, and year? Where are you? Who am I?”

Tremor: Arms extended. Reach for object. Optional: walk across hall

0 - No sweating visible
1 - Palms moderately moist
2 - Beads of sweat visible on forehead
0 - No hallucinations
1 - Tactile hallucinations only
2 - Visual and/or auditory hallucinations
0 - Oriented
1 - Disoriented for date by 1 month or more
2 - Disoriented to place or person
0 - No tremor
1 - Minimally visible tremor
2 - Mild
3 - Moderate
4 - Severe

Figure A.1: SHOT protocol [16].
## Appendix A. Medical Protocols

### Figure A.2: CIWA-Ar protocol [16].

<table>
<thead>
<tr>
<th>NAUSEA AND VOMITING</th>
<th>AGITATION</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ask “Do you feel sick to your stomach? Have you vomited?”</td>
<td><strong>Observation</strong></td>
</tr>
<tr>
<td><strong>Observation</strong></td>
<td>0 normal activity</td>
</tr>
<tr>
<td>0 no nausea and no vomiting</td>
<td>1 somewhat more than normal activity</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>4 moderately fidgety and restless</td>
</tr>
<tr>
<td>4 intermittent nausea with dry heaves</td>
<td>5</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>6</td>
<td>7 paces back and forth during most of the interview, or constantly thrashes about</td>
</tr>
<tr>
<td>7 constant nausea, frequent dry heaves and vomiting</td>
<td></td>
</tr>
<tr>
<td><strong>TREMOR</strong></td>
<td></td>
</tr>
<tr>
<td>Arms extended and fingers spread apart</td>
<td></td>
</tr>
<tr>
<td><strong>Observation</strong></td>
<td></td>
</tr>
<tr>
<td>0 no tremor</td>
<td></td>
</tr>
<tr>
<td>1 not visible, but can be felt fingertip to fingertip</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
<tr>
<td>4 moderate, with patient’s arms extended</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
</tr>
<tr>
<td>7 severe, even with arms not extended</td>
<td></td>
</tr>
<tr>
<td><strong>PAROXYSMAL SWEATS</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Observation</strong></td>
<td></td>
</tr>
<tr>
<td>0 no sweat visible</td>
<td></td>
</tr>
<tr>
<td>1 barely perceptible sweating, palms moist</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
<tr>
<td>4 beads of sweat obvious on forehead</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
</tr>
<tr>
<td>7 drenching sweat</td>
<td></td>
</tr>
<tr>
<td><strong>ANXIETY</strong></td>
<td></td>
</tr>
<tr>
<td>Ask “Do you feel nervous?”</td>
<td></td>
</tr>
<tr>
<td><strong>Observation</strong></td>
<td></td>
</tr>
<tr>
<td>0 no anxiety, at ease</td>
<td></td>
</tr>
<tr>
<td>1 mildly anxious</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
<tr>
<td>4 moderately anxious, or guarded, so anxiety is inferred</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
</tr>
<tr>
<td>7 equivalent to acute panic states as seen in severe delirium or acute schizophrenic reactions</td>
<td></td>
</tr>
<tr>
<td><strong>HEADACHE, FULLNESS IN HEAD</strong></td>
<td></td>
</tr>
<tr>
<td>Ask “Does your head feel different? Does it feel like there is a band around your head?”</td>
<td></td>
</tr>
<tr>
<td>Otherwise, rate severity.</td>
<td></td>
</tr>
<tr>
<td><strong>Observation</strong></td>
<td></td>
</tr>
<tr>
<td>0 not present</td>
<td></td>
</tr>
<tr>
<td>1 very mild</td>
<td></td>
</tr>
<tr>
<td>2 mild</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
<tr>
<td>4 moderately severe</td>
<td></td>
</tr>
<tr>
<td>5 severe</td>
<td></td>
</tr>
<tr>
<td>6 very severe</td>
<td></td>
</tr>
<tr>
<td>7 extremely severe</td>
<td></td>
</tr>
</tbody>
</table>

Total CIWA-A score: __________

A CIWA score of 10 or more indicates the need for benzodiazepine treatment.

CIWA = Clinical Institute Withdrawal Assessment.