Artifact Detection in Physiological Parameter Trend Data

By

Kuo-Hsiung Hanson Wong

Submitted to the Department of Electrical Engineering and Computer Science in Partial Fulfillment of the Requirements for the Degree of Master of Engineering in Electrical Engineering and Computer Science at the Massachusetts Institute of Technology

February 10, 2003

Copyright 2003 MIT. All rights reserved.
Artifact Detection in Physiological
Parameter Trend Data

By
Kuo-Hsiung Hanson Wong

Submitted to the Department of Electrical Engineering and Computer Science
February 10, 2003

In Partial Fulfillment of the Requirements for the Degree of
Master of Engineering in Electrical Engineering and Computer Science

Abstract
Physiological signals recorded in an intensive care unit are often corrupted by artifactual
data. This results not only in false alarms but also in problems analyzing the data in a
research environment. This project presents an alternative method for artifact detection
using both higher-level limits on single-signal variation as well as correlations between
multiple related signals. Analysis of the algorithm was performed using parameter trend
data from 34 ICU patients. Problems with this specific data set resulted in lower values
for sensitivity and positive predictivity. Overall, the sensitivity and positive predictivity
for the algorithm are 78% and 66%, respectively.

Thesis Supervisor: Roger G. Mark, M.D., Ph.D.
Distinguished Professor in Health Sciences and Technology, HST
Professor of Electrical Engineering and Bioengineering,
1.1 ICU

Patients in an intensive care unit (ICU) are monitored using devices that measure and record a large number of physiological signals with the purpose of obtaining a comprehensive view of their physiological state at a particular time. These devices usually incorporate simple alarm functions that are meant to alert the ICU staff to the arrival of an "event," a change in the signal property that is associated with an underlying pathophysiological process. Such alarms make use of threshold detection and sound when a signal value exceeds a certain threshold. Signals of interest often include heart rate, blood pressure, and other variables but differ depending upon the patient and his or her condition.

Comparisons with predetermined thresholds are done on each signal separately with no inter-signal relationship taken into account. Problems arise, in part, from this lack of coupling. Certain physiological signals are correlated with each other, as is the case with heart rate and mean arterial blood pressure. However, if one single-signal alarm is triggered, other signals are not examined to see if similar events are detected among them. A result of this process is a prevalence of false alarms.

An alarm goes off every 30 seconds in a typical full ICU. Studies have shown that a high percentage of these alarms are incorrect. Zong found that 35% of alarms in a collection of ICU recordings were false while Tsien found that up to 80-90% of all alarms could be classified as not meaningful. Though a small percentage (6%) of alarms in Tsien’s study were clinically irrelevant true alarms, the remaining were false-positives. The effect of this large number of false alarms is not only unnecessary stress
placed on patients and their families but also a reduction in the trust placed on the machines by the ICU staff.

Artifacts - significant changes in the values of the signal not due to physiological changes in the patients - are common in ICU data recordings and are the predominant cause of false alarms. Such artifacts may also result in delayed warnings or, in the worst case, prevent the identification of important events altogether. Unfortunately, there are many causes of artifacts. Interference, unreliable transducers, problems with the connection to the monitor, and problems with calibration of monitoring devices are all potential sources of artifacts. In addition, patient movement can also lead to artifactual recordings.  

The following figures illustrate the effect of artifacts on alarms. All figures show blood pressure waveform recordings from a single ICU patient. Figure 1.1 shows a recording with no instances of artifactual data. The high heart rate (~120bpm), high estimated respiratory rate (~40/min), and other features in this signal may be indicative of a patient with pneumonia or pulmonary edema. However, the values are within normal bounds as determined by the bedside monitor, and no alarm is triggered during this period. Figure 1.2 similarly shows a recording of the blood pressure waveform lacking artifacts. The signal values during this period are low, and the monitor records a mean blood pressure value of 75mmHg. This is below the user-defined threshold of 80mmHg, and an alarm is appropriately sounded at this time. Figure 1.3 shows a corrupted segment of data. Though the blood pressure at the beginning of this period is still low, it is above the monitor threshold and does not trigger an alarm. However, the monitor is unable to
differentiate between the true data and the artifact. The result is a monitor reading of 5mmHg, well below the threshold, and an alarm is triggered - a false alarm.

Figure 1.1 shows a blood pressure recording from an intensive care unit patient. Notice the innate variability in maximum and minimum values in the signal due to respiration.
Figure 1.2 shows a blood pressure recording from the same patient that has gone below the minimum threshold. An alarm is sounded at this time.

Figure 1.3 shows a blood pressure recording from the same patient that has gone below the minimum threshold. Notice that the signal during this time contains an artifact and leads to a false alarm.
We are able to see from these figures an example of the noise that exists in physiological recordings. A short summary on resolving false alarms will be given in the first section of this chapter. Following this review, a context for further research will be presented. All this information will lead up to the hypothesis and goal of this thesis: to create an improved algorithm for artifact detection.

1.2 Resolving False Alarms

The simplest method of resolving false alarms is to ignore them. Though this ethically may not be a valid option, the practice is unfortunately partially implemented as staff in the ICU becomes desensitized to the large number of false alarms. This "crying wolf" effect may result in potential delays in response or in complete ignoring of the alarms. Altering the thresholds - increasing the range of "normal" values - can also decrease the rate of false alarms, but this would in turn decrease the number of true-positives.

More applicable methods have focused on addressing the quality of the underlying physiological data. Decision support systems, such as the monitors in the ICU, can only be as precise as the data they are based on. A need thus arises for artifact detection and/or removal. Two prevalent forms currently in practice are rule-based filtering and median filtering.

1.2.1 Rule-Based Filtering

Rule-based filtering relies on a set of conditions under which data can be tagged, or labeled, as artifactual. The most common conditions establish physical boundaries beyond which data is deemed to be physiologically impossible. Limits can also be set on
the acceptable range of variability from one data point to another. The latter can be done by monitoring either the standard deviation of a signal over a certain period of time or the difference in value of the signal from one time point to the next. Both boundary and variation limits require an understanding of the physiological signals themselves in order to create reasonable thresholds. Correlation rules can also be included to determine the validity of a signal by marking a certain time segment of a signal as noisy if another related signal is found to contain artifacts in the same time period.\textsuperscript{7}

Difficulties with rule-based filters arise especially when handling data from patients in the ICU. "Normal" thresholds are not uniform among all patients, and critically ill patients may have stable readings that are abnormal for a healthy patient.\textsuperscript{8} As such, any limits must be broad enough to account for possible physiological variation, which in turn may unfortunately compromise the ability to detect actual artifacts. Figure 1.4 provides a sample trend recording and the output after processing by a rule-based filter.
Figure 1.4 (top) shows a trend recording of central venous pressure measured every minute. Figure (bottom) shows the results of a rule-based filter with a maximum boundary limit of 80mmHg, minimum boundary limit of 0mmHg, and a maximum deviation limit of 5mmHg over a period of one minute.

1.2.2 Median Filtering

Median filtering is a non-linear signal-processing algorithm that examines each data point and a "window" of points around it. The median for the set of data within the window is calculated and substituted for the actual value. The median filter is useful in treating extreme outliers, in contrast with a mean filter, due to its zero impulse response. However, longer lasting artifacts are often not properly suppressed. This can be addressed by increasing the window size, though increasing the length of the window
decreases the incidence of both false and true alarms. In addition, although such filtering may be useful in critical settings, the resulting loss of information makes the technique less appropriate for domains that are characterized by relatively sparse data. Nevertheless, median filtering remains a predominant form of artifact removal in use today.

Figure 1.5 (top) again shows the previous recording of a central venous pressure trend. Figure (middle) shows the results of passing the data through a median filter with a window size of three minutes, which here equates to three data points. Figure (bottom) shows the results of using a median filter with a window size of five data points. Notice how the "spike" noise is filtered out as the window size is increased but the "step" noise is not. Notice also the loss of data as the larger window smooths out the recording.
Figure 1.5 alludes to a larger problem associated with median filters. As with rule-based filtering, an understanding of any signal of interest is needed in order to determine the optimum window size for use. Signals that vary significantly over short periods of time may lose necessary degrees of detail as they are processed by the filter. Such data, as shown in Figure 1.5 (bottom), will be lost as they are smoothed out. Physiological recordings such as CVP, especially from patients in an ICU, are apt to vary suddenly. Loss of such finer details from filtering can prove costly.

1.2.3 Intelligent Alarms

Efforts are underway to develop more "intelligent" monitors that may be able to discern false alarms. Such instruments would incorporate algorithms able to monitor trends in physiological signals instead of relying only on instantaneous values. Incorporating both trends and physiological models, multiparameter methods are being devised to allow for some predictive power as well as the ability to discern noise in a signal data stream. Nevertheless, before signals can be passed on for interpretation, it is important to assess the quality of the signal itself. As monitors are being improved and more signal processing is being applied to the data, it is necessary to likewise continue to improve noise detection in order to better understand the validity of the underlying data.

1.3 MIMIC

Artifacts play a role not only in a clinical setting but also in medical research. Certain experiments, such as the development and testing of medical decision support systems, require large amounts of well-characterized test data. The MIMIC (Multiparameter
Intelligent Monitoring for Intensive Care) database was created to meet those needs. The database is an archive of patient records including waveform signals and vital sign numeric trends collected from bedside monitors along with nurses' progress notes, laboratory results, medication profiles, and other forms of clinical data. These are taken from patients in the medical, surgical, and cardiac intensive care units of a partner hospital. Work has been done on improving and expanding MIMIC, and it now contains data from thousands of patients.\textsuperscript{11,12}

Working with data from the MIMIC database, multiparameter trend monitoring using wavelet analysis has shown promising results. The trend data consists of measures of the physiological waveforms taken once every minute, and it is analyzed using wavelets to detect patterns that may lead to the diagnosis of certain events.\textsuperscript{13} As with any data recording, problems arise with the records in MIMIC due to noise artifacts. While computers may eventually be able to satisfactorily detect and remove such noise, at present only manual removal is possible for the majority of cases.

1.4 Thesis Scope and Goals
The major purpose of the research to be described in this paper is to identify methods of examining physiological data and determining its validity by locating artifacts. The goal is to classify individual areas of a signal in terms of their ability to represent the underlying physiological state of the patient, and in this way assign a level of user confidence in the data. More specifically, as shown in Figure 1.6, the objective is to annotate regions of physiological data as noisy, uncertain, or clean. This work will build
upon current rules-based systems and include analysis of correlation between physiological signals.

Figure 1.6 shows a block diagram of the proposed artifact detector.

This project coincides with efforts to use the MIMIC database to develop wavelet algorithms in discerning trends for use in intelligent patient monitoring. Moreover, as was mentioned previously, knowledge of the underlying signal quality is important in any type of patient monitoring system. The aim of this project is thus to aid subsequent analysis of physiological data by labeling regions so that following stages can then choose to ignore or accept regions based upon their "tag."

However, as this research is concerned specifically with data used for the wavelet trend analysis, actual waveform data will not be discussed. Trend data, as discussed in this paper, is defined as signal recordings measured once every minute. Though the focus
of this paper will be on the research applications of artifact detection with an emphasis on such trend data, any algorithms presented may also be applicable in a clinical setting or other instances where waveform data is used. In addition, the research described here is concerned mainly with artifact detection. While the actual removal or replacement of artifactual data will be briefly discussed, such research is beyond the scope of this paper.

The following chapter will provide a brief background into the physiological signals involved in this research. A method is then presented to examine these signals and their validity, and results with analysis will follow.
2.1 Signal Selection

A large number of physiological signals are recorded from patients during their stay in the ICU. As these recordings often last for periods ranging from many hours to days, analysis of each signal for all patients would be a daunting task. The first stage of this project was thus to select a subset of signals to be used in the development and testing of any potential algorithm. This selection process was followed by a review of numerous trend recordings of each selected signal in an effort to become acquainted with the many different types of artifacts possible in physiological data.

Signals were chosen in large part due to their value in discerning information about the underlying physiology of the patient. The research described here seeks to build upon current artifact detection algorithms by incorporating the interrelation of signals, and thus signals were also chosen such that there was a high degree of correlation among them. Additional factors considered included the amount of patient data available for the respective signals as well as their inherent level of noise.

Four signals were initially chosen that adequately fulfilled these criteria: heart rate; arterial blood pressure; central venous pressure; diastolic pulmonary artery pressure. Though this list was later narrowed to include only heart rate and blood pressure, all four signals are of great value. It is hoped that this project can be extended in the future, beginning most likely by reexamining central venous pressure and diastolic pulmonary artery pressure. A brief description of each variable is thus given subsequently in this chapter.
2.2 Heart Rate

A patient’s heart rate relates the frequency at which his heart beats and is measured in units of beats per minute (bpm). A normal adult has an average resting heart rate of approximately 70 bpm, though heart rate may rise as high as 180 bpm during exercise.\textsuperscript{14} Certain people, such as athletes, will experience lower heart rates as their hearts are able to pump a larger volume of blood during every beat. This volume is called the stroke volume.

The body can regulate the cardiac output - the amount of blood pumped by the heart each minute - by increasing or decreasing heart rate. The relationship between cardiac output and heart rate can be defined as follows:

\[
\text{(Cardiac Output)} = (\text{Heart Rate}) \times (\text{Stroke Volume})
\]

This equation is useful in understanding the effect of heart rate on blood pressure, a topic discussed later in section 2.3. In addition to heart rate, myocardial contractility, preload, and afterload affect the cardiac output. Though preload and afterload are not considered in this thesis, preload is reflected in CVP and DPAP recordings and will be discussed in brief at the end of this chapter.

A typical heart rate trend can include large "spikes" where the signal value may increase dramatically in a short amount of time. An example is shown in Figure 2.1. These sudden jumps may be attributable, for example, to either a double-counting in the ECG waveform where one QRS complex is misidentified as two, or to a sudden heart rhythm change. The former results in an artifactual heart rate value and does not reflect
the true state of the patient. However, as alluded to previously, this study will be performed using data collected from ICU patients. The task of identifying artifacts is more difficult with such data as seemingly artifactual regions may be due to abnormal conditions present in ICU patients. During these instances such spikes may be attributed to an actual underlying physiological condition. One cause for such an event in the heart rate recording could be atrial fibrillation.

![Heart Rate Trend Data](image)

Figure 2.1 shows a typical heart rate trend pattern from an ICU recording.

Atrial fibrillation, the most common abnormal heart rhythm, is an arrhythmia that occurs in various types of chronic heart disease. Under this condition, the atria, or upper chambers of the heart, do not contract and relax sequentially. The atria instead undergo a continuous, uncoordinated, rippling motion and are no longer in sync with the lower chambers of the heart. No constant interval occurs between successive QRS complexes resulting in an irregular heart rhythm that can reach up to 160 bpm.\textsuperscript{14,15}
Any potential artifact detector must attempt to distinguish between these true physiological events from artifacts. This can be accomplished, in part, by examining the activity of other signals closely associated with heart rate. A true event that alters the heart rate trend data should also be evident elsewhere in related physiological signals. Blood pressure has a high level of correlation with heart rate and is one such signal that can be used to cross-check uncertain variations.

2.3 Arterial Blood Pressure

The arterial blood pressure (ABP) pulse is generated by flow from the heart into the ascending aorta. Ejection of blood into the aorta dilates the aorta and generates a pressure wave. This wave is then propagated to other arteries throughout the body. Figure 2.2 is shown again from section 1.1 to provide an example of a blood pressure waveform recording.
Figure 2.2 shows a blood pressure waveform from an ICU recording.

The maximum value during each cycle is the systolic blood pressure whereas the minimum value is the diastolic blood pressure. The difference between the systolic and diastolic ABPs is defined as the pulse pressure. The average value over one cycle is the mean blood pressure. This value is a running average calculated independently from the systolic and diastolic measurements. A typical value for systolic ABP in a resting adult is 120 mmHg, 80 mmHg for diastolic ABP, and 90 mmHg for mean ABP.

The mean blood pressure and flow in the cardiovascular system follow a form of Ohms law, namely

\[
MABP = (CO) (R) \quad 2.2
\]

where cardiac output (CO) is as defined by equation 2.1 in section 2.2 and R is the peripheral resistance. While frictional resistance is relatively small in the large arteries,
small arteries offer moderate resistance to blood flow. This resistance reaches a maximum in the arterioles. The pressure drop is greatest here across the terminal segment of the small arteries and the arterioles. The body is able to control the value of R by adjusting the degree of contraction of the circular muscle of these small vessels. This not only permits regulation of tissue blood flow but also aids in the control of arterial blood pressure.\textsuperscript{14} Equation 2.2 applies to both the systemic and pulmonary circulations.

Substituting for CO in the previous equation we arrive at

\[
\text{MABP} = (HR \times SV) \times R
\]

2.3

Here, one begins to see the relationship between heart rate and blood pressure. This equation is applied, for example, as the cardiovascular system attempts to keep MABP constant using certain feedback mechanisms throughout the body. Changes in arterial blood pressure initiate a reflex that leads to an inverse change in heart rate and peripheral resistance. If MABP is lowered, it follows from equation 2.3 that raising any combination of HR, SV, and R, will cause a normalizing effect on MABP. This reflex is initiated, in part, by the baroreceptors.

The baroreceptors are stretch receptors located in the carotid sinuses and in the aortic arch. The pressoreceptor nerve terminals in the walls of the carotid sinus and aortic arch respond to the stretch and deformation of the vessel induced by the arterial pressure. With an increase in blood pressure, the frequency of impulses arising from the receptors also increases. The brainstem responds by decreasing sympathetic outflow while increasing parasympathetic outflow. The decreased sympathetic tone results in vasodilation and a drop in cardiac contractility. The resulting lowered heart rate is
reduced further by the changes in parasympathetic tone. Blood pressure therefore decreases due to the decrease in both cardiac output and in peripheral resistance in accordance with equation 2.2. The arterial baroreceptors thus play a key role in short-term adjustments of blood pressure in response to relatively abrupt changes in blood volume, cardiac output, or peripheral resistance.\textsuperscript{14}

The role baroreceptors play in the regulation of blood pressure can be seen in the body's response to a hemorrhage. An individual who has lost a large quantity of blood experiences a weak arterial pulse as the arterial systolic, diastolic, and pulse pressures decrease. The reduction in MABP and in pulse pressure during a hemorrhage decreases the stimulation of the baroreceptors. As a result, several cardiovascular responses are evoked. Heart rate is increased and is accompanied by a related venoconstriction and arteriolar constriction. The resulting increase in peripheral resistance further minimizes the fall in arterial pressure. Figure 2.3 provides an illustration of the effect of baroreceptors after a controlled blood loss.\textsuperscript{14}
Figure 2.3 shows the effect of the baroceptors after an 8% blood loss in three groups of dogs. (A) shows the resulting blood pressure decrease in dogs where the aortic reflexes were interrupted. (B) shows the resulting blood pressure decrease in dogs where the carotid sinus reflexes were interrupted. (C) shows the resulting blood pressure decrease in dogs where all baroreflexes were interrupted.

This response to hemorrhage further demonstrates the relationship between MABP and HR as described in equation 2.3. It should be noted, though, that the baroreceptor reflex is a feedback mechanism, and there will thus be a necessary corresponding error term associated with the blood pressure. Though heart rate and peripheral resistance result in an increased blood pressure, blood pressure does not return exactly to its pre-hemorrhage level. The net effect, factoring in the initial drop in blood pressure due to the blood loss, is a drop in blood pressure accompanied by an increase in heart rate and peripheral resistance.

Nevertheless, holding other variables constant, if HR increases and there is no corresponding decrease in stroke volume, cardiac output and thus MABP should experience a similar increase in value. Such an event may occur during a sharp sensation of pain or other outside stimulus that can lead to a rise in heart rate. The relationship
between HR and MABP can be useful in determining the validity of data such as that in Figure 2.1. Following from equation 2.3, the large increase in the heart rate recording, if it accurately portrays the patient’s state, should also be evident, in some form, in the blood pressure recording. Unfortunately, there are instances where blood pressure will fall even as heart rate increases, and other physiological data must also be taken into account.

A sudden rise in heart rate may be due to an onset of tachycardia, or an increase in the contraction frequency of the heart. In the case of supraventricular tachycardia, heart rate can suddenly reach up to 140 to 250 bpm. Such rapid contractions may not allow sufficient time for ventricular filling, and the heart is no longer able to function properly. Blood is no longer effectively being pumped into the body, and such situations can be accompanied by an exponential drop in blood pressure towards Pms - the mean systemic filling pressure. This is the pressure if there were no flow in the circulatory system. The difficulty in detecting artifacts lies in the fact that there are many causes for events, whether true or false in nature. Using the relationship between signals is helpful in determining the validity of a particular data set, but the complex interworkings of heart rate and blood pressure help illustrate that a thorough understanding of these relationships is needed. Further compounding the problem is the fact that associated variations, in and of themselves, do not necessarily indicate true and uncorrupted data.

Certain cases of artifactual noise pertain to a single signal. For example, a faulty electrode/lead will cause false heart rate recordings while the other signals will be unaffected. However, other situations may affect multiple channels. A patient coughing or moving may shift both the heart rate as well as the blood pressure recording though the
patient’s underlying physiology remains unchanged. Again, any potential artifact
detector must attempt to distinguish between these and true physiological events.

Lastly, just as heart rate and blood pressure are related, the individual components
of blood pressure are also interrelated. Systolic, diastolic, and mean ABPs are positively
correlated with each other as a change in one signal is usually accompanied by a similar
change in the remaining two. Though these changes may not be of the same magnitude,
an increase in the MABP should result in an increase in both the systolic and diastolic
ABP, as well. If the stroke volume is increased by a factor of two while HR and R are
held constant, equation 2.2 shows that MABP should similarly double in value. Under
constant compliance, this increase in stroke volume will increase systolic and diastolic
ABPs, but the systolic pressure will increase more so such that the pulse pressure will be
approximately twice as great as before. Figure 2.4 shows a patient trend recording for
heart rate and the individual blood pressure measurements.
Figure 2.4 illustrates the relationship between heart rate and blood pressure. In general, variations in the trend data shown here are evident in both the heart rate and blood pressure recordings. Notice, also, that there is a much higher degree of correlation between the systolic, mean, and diastolic blood pressures than between the heart rate and blood pressure.

2.4 Central Venous Pressure (CVP)

Central venous pressure is the pressure in the right atrium and thoracic venae cavae and indicates the pressure of the blood as it returns to the right side of the heart. An approximation of right ventricular end diastolic pressure, or right ventricular preload, CVP reflects right ventricular function. On average, CVP is around 1.5 mmHg but can range from just above 0mmHg to 5mmHg. The recording shown in Figure 1.4 is thus unusually high, though that is indicative of many patients in an ICU.
CVP is closely linked to blood pressure in that it also affects cardiac output. This relationship, however, is defined by two functions - the cardiac function curve and the vascular function curve. The cardiac function curve describes the effect of venous return on cardiac output and depends upon the characteristics of the heart. The vascular function curve describes the effect of cardiac output on venous return and depends upon the characteristics of the vascular system. Though beyond the scope of this thesis, the actual value of CVP and cardiac output depend upon the intersection of these two curves.

Figure 2.5 shows a central venous pressure trend pattern from an ICU recording alongside other recordings of interest.

2.5 Diastolic Pulmonary Artery Pressure (DPAP)
Diastolic pulmonary artery pressure is used to approximate the value of filling pressure of the left heart, or the end-diastolic pressure. Similar to CVP and its relationship to the right ventricle, DPAP is a measure of left ventricular preload. A normal value of DPAP for resting adults is around 9 mmHg.

Figure 2.5 shows a diastolic pulmonary artery pressure trend pattern from an ICU recording alongside other recordings of interest.

2.6 Summary
The physiological signals presented here are interrelated with each other. These relationships, if properly understood, can be used to enhance artifact detection in each of the signals. Though figure 2.5 illustrates numerous trend recordings for an ICU patient, henceforth the focus will remain solely on heart rate and blood pressure signals.
Figure 2.5 shows the various trend data recordings for a particular individual. Note how variations in certain signals are evident in other signals, as well. Note also how there are instances where such a relationship is not clearly seen.
Chapter 3 - Artifact Detector Algorithm

3.1 Algorithm Overview

Many different types of artifacts exist in trend data, and only a few examples have been presented thus far. As shown in the previous chapters, certain instances of artifacts are simple to identify while others are not as easily detected. This wide variety of noise presents a problem in that any algorithm used to detect artifacts must be robust enough to recognize a large proportion of noise yet be fine-tuned enough to distinguish true events caused by a patient’s underlying physiological state. The goal of the algorithm to be described here in this thesis to "tag" each region of physiological trend data from patients contained in the MIMIC database as one of three categories: reliable and clean of artifacts; uncertain; definitely corrupted by noise.

The algorithm has been, to this point, presented as a black box. This chapter will discuss the artifact detector shown in Figure 1.6 in more detail. Figure 3.1 shows the artifact detector separated into its four components: missing data detector; threshold detector; variation detector; correlation detector. The following sections of this chapter will each discuss various types of artifacts, describe a particular subsection of the algorithm, and explain how it is designed to identify such artifacts.

In general, save for the missing data detector, each subsection is distinct from one another. The outputs from all detectors consist of regions where data has been tagged as uncertain and regions where data has been tagged as corrupted. These tags are then combined to create an aggregate set of uncertain and corrupted locations. Regions of data not tagged as either uncertain or corrupted are tagged as clean.
Figure 3.1 expands figure 1.6 and depicts the aggregate artifact detector with its subsections: missing data detector; threshold detector; variation detector; correlation detector.

3.2 Missing Data Detector

Missing data in a recording is an easily identifiable class of artifact, and this lack of data may be attributed to any number of causes. A nurse may have disconnected the probe for the respective signal or there may have been a bad connection either to the patient or to the monitor. The function of the missing data detector in this project is twofold - it detects and stores locations of missing data and then "fills in" those regions.

Periods where no data is recorded are displayed differently based upon the devices being used to measure and collect the signals. The method to be described here is unique to the patient trend data stored in the MIMIC database. However, similar methods can be designed for different record formats if the manner of recording missing data is known. Regions of missing data in the MIMIC database are assigned the value of
Such a value is far beyond the possible range of any physiological signal and is thus easy to detect. Figure 3.2 provides an example of a recording with missing data.

Figure 3.2 shows a heart rate trend recording containing missing data. Notice the large section of missing data in the beginning compared to the smaller segments throughout the remainder of the recording. Figure (bottom) shows an expanded region from the recording. The regions of missing data differ greatly from those where data exists.

The detector locates regions of missing data by searching for all points in the trend recording of the value -888. Each instance is tagged as corrupted, and its time stamp is saved and passed on to be combined with artifact locations as determined by the other subsections of the algorithm. The missing data detector then replaces each instance of a -888 value in the trend data with an estimated value.
The question arises as to whether or not areas with missing data should be "filled in" and, if so, by what process should this be done. This study has chosen to use substitute values in order to provide usable data for calculations in subsequent parts of the artifact detector. Certain stages following the missing data detector perform calculations on data over a window of time. While large sections of missing data (specifically those much larger than that of the window size) will not be worthwhile to pass through to subsequent detectors, the algorithm currently does not limit the size of a missing data region to be filled in. However, the substitution process is most beneficial when smaller portions of data, those proportional in length to the window size, are missing. In such cases, as in the absence of a single data point, calculations involving the area of noise and its surrounding data may not be possible, thus leading to the entire area within the window to be declared artifactual.

A "best guess" for the missing data is thus used, and this estimate is determined using linear interpolation. The missing data detector takes note of the last true data point that is recorded before the section in question and the first true data point recorded afterwards. These two endpoints are used as a basis for linear interpolation, and the missing data is replaced with values from the resulting line between these two points. Instances of missing data that occur either at the beginning or end of a patient record do not provide a suitable beginning or end point for use in such a calculation. During these cases, a NaN value is substituted for the missing data, and no estimated values are passed on to following sections of the algorithm. The NaN value indicates in the mathematical program used for this thesis that there is “not a number” at that location. Further calculations around this region will not be possible.
Limitations to this method arise as physiological signals are inherently nonlinear. Linear interpolation thus, in and of itself, produces artifacts. Nevertheless, the resulting artifacts should be minimal in the short-time-period situations where this method is most beneficial. Future work may attempt to study additional methods for managing missing data. The process for missing data detection is illustrated in figure 3.3.

![Block diagram of the missing data detector](image)

Figure 3.3 shows a block diagram of the missing data detector. Though heart rate is not displayed here, it is filtered using the same methodology.

3.3 Threshold Detector

The preceding section introduced the situation where a physiological signal exhibited extremely low values, notably that of -888 during regions of missing data. Other values - those that are less inordinate but likewise beyond the range of feasibility - can also appear
in physiological trend data. More common in recordings are values that go above possible bounds, as shown in the sudden large step increase of CVP in figures 1.4 and 1.5. However, as CVP values are normally near zero, shifts in body position that can affect the signal’s baseline may result in negative CVP values. The reliability of such a data point is highly questionable. These artifacts, where values exceed either the maximum or fall below the minimum thresholds of physical limitation, are, as with missing data, simple to detect.

The threshold detector receives as inputs two parameters in addition to the trend data to be analyzed. The user must specify a maximum value and a minimum value with relation to each signal. These values may be a constant value or that of another signal. Figure 3.4 provides an example of a blood pressure trend recording passed through the threshold detector as well as the resulting output tags.
Figure 3.4 (top) is a blood pressure recording showing systolic, mean, and diastolic blood pressures. The remaining sections show the result of passing these through the threshold filter. Note how regions where, for example, the systolic blood pressure falls below the mean blood pressure are tagged as artifactual by both the systolic and mean detectors.

The mean pressure, by definition, cannot exceed the systolic blood pressure or fall below the diastolic blood pressure. However, as the mean blood pressure is calculated independently of systolic and diastolic measurements, an artifact in the systolic blood pressure signal may not appear in that of mean blood pressure. Though certain physiological conditions result in all three values being of similar magnitude, it is reasonable to set the maximum threshold for the mean blood pressure to be the systolic blood pressure and the minimum threshold to be the diastolic blood pressure. In addition, constant thresholds are used to evaluate the mean blood pressure. These set values are
used to address situations where either the systolic or diastolic blood pressure is
artifactual and thus cannot provide a good basis with which to check the mean blood
pressure.

The threshold detector locates all regions where the mean blood pressure exceeds
either the systolic blood pressure or the maximum constant threshold, and it likewise
locates regions where it falls below either the diastolic blood pressure or the minimum
constant threshold. A similar process is applied to the systolic and diastolic blood
pressures, and all time stamps of regions where values exceed their associated limits are
tagged as corrupted. This set of locations is later combined with those similarly tagged
by other detectors. A diagram of this process is provided in figure 3.5.

The difficulty with the threshold detector is the same as that of other rule-based
filters described in chapter 1. Any maximum threshold must be set high enough to allow
for actual physiological readings that are abnormally high but low enough to filter out
areas that are not physiologically possible. Determining thresholds requires a thorough
understanding of the signals themselves. Consequently, the values used in this project
were chosen after a review of the literature, discussion with experts, and analysis of
sample data.
Figure 3.5 shows a block diagram of the threshold detector as applied to blood pressure.

3.4 Variation Detector

Sections 3.2 and 3.3 presented two methods for detecting artifacts. However, a large number of artifacts that are not encompassed within the missing data category lie within reasonable physiological bounds. Certain such artifacts can be detected in a manner similar to that used by the threshold detector. Just as a blood pressure trend recording at any point in time cannot exceed a certain value, the change in blood pressure from one value to the next also cannot exceed a certain value. Though the mean blood pressure of a patient in an ICU can rise above 250 mmHg, it is not likely to do so if its value a minute earlier was below 100 mmHg. These spikes in value can be common in physiological recordings, and more advanced methods than the ones previously described are needed to detect such artifacts.
The variation detector searches for regions where there may be unreasonable changes in signal values by performing a moving standard deviation calculation over a set of trend data. Similar to the median filter described in chapter 1, the detector calculates the standard deviation for each point and a "window" of points around it. This window slides over the entire data set until the standard deviation around each point is calculated. The result is a data set of standard deviations, one for each point in the corresponding trend data recording. This new set of standard deviation data will henceforth be referred to as movstd. The overall standard deviation and mean of movstd are then calculated.

Figure 3.6 shows a block diagram of the variation detector. Though heart rate is not displayed here, it is filtered using the same methodology.
The variation detector, as shown in figure 3.6, receives three sets of parameters in addition to the trend data to be analyzed. The user must first specify a window size, var_win, for the moving standard deviation calculations. This input may be a vector containing multiple values, thus allowing for the detector to perform variation analysis over more than one window size. Larger windows allow for analysis of slower variations in the signal that may not be detected with smaller windows. Determination of window size is discussed in section 4.4.

Var_limmul, a user-defined constant, is then used to calculate a limit for determining regions of excess variation. The threshold for each signal is calculated using the following formula

\[ \text{threshold-variable} = (\text{standard deviation of movstd}) \times \text{var_limmul} + (\text{mean of movstd}) \]

where movstd = moving standard deviation calculation of signal in question

and var_limmul = user-defined constant

The standard deviation of movstd is multiplied by var_limmul, and this value is added to the mean of movstd. The resulting threshold-variable is compared against each value of movstd, with those values exceeding it being tagged as artifactual. In this way, the variation detector searches for regions of unusual deviation, where the standard for variation is determined in relation to each individual patient. Statistical variables, specifically those of standard deviation and mean, are unique to a specific patient and thus allow the detector to define variation thresholds better suited to each patient.

Var_limcon, a maximum threshold value, is used to address situations where there may be extreme outliers in the data set. These outliers may raise the standard deviation
of movstd so high such that the value of threshold-variable is likewise very high. In these
instances no regions besides that of the outliers are tagged. The actual threshold for
variation is thus the minimum of threshold-variable and the var_limcon. Both
var_limmul and var_limcon differ for heart rate and for blood pressure.

Figures 3.7 and 3.8 show the calculations and output of the variation detector.

A final step is used to further verify the blood pressure data. As was noted in
equation 2.3, blood pressure is correlated to heart rate. Any large or sudden change in
blood pressure should somehow also be reflected in the heart rate. In the case of
ventricular fibrillation, when blood pressure drops exponentially within seconds, there is
no coherent contraction of the heart and thus no ejection of blood. A variation in the
heart rate is thus accompanied by a variation in the blood pressure.

Regions of blood pressure that have been tagged by the variation detector are thus
checked with the corresponding regions in the heart rate trend data. If the variation
detector has likewise tagged the heart rate, then the region is labeled as uncertain. The
region is not labeled as clear due to the fact that, as described in section 2.3, associated
variations do not in and of themselves fully indicate that the recordings are uncorrupted.
Regions with no corresponding tag in the heart rate recording are labeled as artifactual,
and the remaining regions are labeled as clear. The time stamps of each set - uncertain,
artifactual, and clear - are passed on and combined with those as tagged by other
detectors.
Figure 3.7 shows the output from the first stage of the variation detector. The blood pressure trend recordings are passed through and the moving standard deviations for each signal are calculated. The solid horizontal line indicates var_limcon, the maximum constant value, and the dashed horizontal line indicates threshold-variable as calculated using equation 3.1.
3.5 Correlation Detector

Further analysis of artifacts requires a deeper understanding of physiology beyond checks against maximum values, minimum values, and maximum variation. This can be done through the incorporation of relationships between physiological signals. Current patient monitors make use of single-signal analysis and, as mentioned in Chapter 1, this has proven to be a less-than-ideal solution. Though the variation detector made use of inter-signal rules to account for relationships among recordings, this was limited to identifying if other signals had likewise been tagged as potentially corrupted. As such, the
verification of blood pressure variation based on heart rate variation was used more as an additional check than as an actual stand-alone detector. More subtle artifacts would pass through this system undetected.

A clot in a catheter with a transducer measuring blood pressure produces a recording such that the mean blood pressure value remains steady while the diastolic and systolic blood pressure values both begin to approach the mean. The clot thus acts as a low pass filter of the blood pressure signal. Figures 3.4 and 3.8 illustrate a potential clot around the 35th hour in the recording. Though this individual recording experiences high variation during this time, as shown in figure 3.8, were the transition of the diastolic and systolic values somewhat smoother, none of the aforementioned detectors would tag this region as artifactual. More robust analysis can be used in such cases involving highly correlated signals.

The correlation detector is the most complex of the subsections of the algorithm and performs a moving cross-correlation calculation over the trends of multiple signals from a single patient. With regards to blood pressure, the correlation detector calculates the correlation between the diastolic blood pressure and mean blood pressure, diastolic blood pressure and systolic blood pressure, and the correlation between mean blood pressure and systolic blood pressure. These values are then examined to detect any changes in the relationship of these three signals.

Similar to the variation detector, the user must specify a window size, cor_win, for the moving correlation calculation. This input may be a vector containing multiple values in order to detect changes over varying amounts of time. Correlations are calculated for each window size, and regions with low degrees of correlation are tagged.
The user must also specify above what percentage of window-size-runs must a region be tagged to be considered artifactual.

Consider an example where the window size input is "[5 10 15]" and the percentage input is ".5". This indicates that window sizes of 5 minutes, 10 minutes, and 15 minutes are to be used in calculating correlation between the blood pressure signals. Suppose a region is tagged during the 5-minute window calculation but is declared clear during the calculations of other window sizes. It has been tagged by less than half of window-size-runs, and the region is thus labeled as uncertain. Suppose another region is tagged during the 5 minute window calculation as well as during the 10 minute window calculation. This region has been tagged by two-thirds of the window-size-runs, greater than the 50% necessary, and this region is thus labeled as artifactual.

Regions of low correlation are determined by examining the absolute value of the difference between the three cross-correlation values. This is done using the following equation:

\[ |\text{correlation of diastolic and mean} - \text{correlation of diastolic and systolic}| + |\text{correlation of diastolic and mean} - \text{correlation of mean and systolic}| + |\text{correlation of mean and systolic} - \text{correlation of diastolic and systolic}| = \text{correlation difference index (CDI)} \]

3.2

This correlation difference index is calculated for each time period in the trend data and increases as the signals become less correlated. The correlation coefficient of each blood pressure signal with one another should normally equal one. In such cases, the difference between correlations would be zero, and the resulting CDI would also equal zero. This does not hold true during instances of corrupted data.
The mean blood pressure is determined independently of both systolic and diastolic measurements. It is thus common for an artifact to exist either only in the mean blood pressure recording, or only in both the systolic and diastolic recording. In such cases, the correlation coefficient between the mean blood pressure and both the systolic and diastolic blood pressures would fall well below one and the CDI value would increase. A similar situation occurs during the clot as described in the beginning of this section. The clot results in reduced correlation coefficients with regards to all three blood pressure signals.

The correlation detector compares each CDI value in a manner similar that used by the variation detector, namely the threshold-value is calculated as follows

\[
(\text{standard deviation of CDI}) \times \text{var}_\text{lim mul}_\text{CDI} + (\text{mean of CDI}) = \text{threshold-variable CDI}
\]

where CDI = correlation difference index as defined by equation 3.2

and \( \text{var}_\text{lim mul}_\text{CDI} \) = user-defined constant

The overall standard deviation of CDI is multiplied with \( \text{var}_\text{lim mul}_\text{CDI} \), a constant specified by the user, and added to the overall mean of CDI. This resulting value is compared against each value of CDI. Those values that exceed either this threshold or an additional user-inputted maximum constant value, var_con_CDl, are tagged. The determination of the tag, whether artifactual or uncertain, is based on the window-size-run percentage as described earlier.

A problem arises during regions of low variability within the signal. Correlation measures the degree of linearity between signals and is highly sensitive during periods when the signal values are relatively constant over a period of time. A small change in
only one such signal will result in a large change in its correlation value with other
signals. Though this may be a normal phenomenon, the large change in correlation will
result in the region being inappropriately tagged as corrupted. The correlation detector
thus filters out regions of low variability and does not tag any data within these regions.
The user specifies a set value for the underlying signal’s standard deviation below which
the correlation detector's output is disregarded.

The correlation detector thus receives five inputs: window size values; window-
size-run-fraction; constant value to be multiplied by the standard deviation of CDI to
calculate threshold-variable_CDl; maximum constant value for CDI variation; minimum
constant value for variation among the underlying signals. Note that the correlation
detector does not attempt to determine which of the three blood pressure recordings
contain artifacts, only that at least one of them is corrupted. Its output is saved and
combined with those of the preceding subsections during the final step of the algorithm.

Figure 3.9 illustrates the correlation detector, while figures 3.10, 3.11, and 3.12
show a sample of patient data processed by the detector.
Figure 3.9 shows a block diagram of the correlation detector.
Figure 3.10 shows the output from the first stage of the correlation detector. Figure (top) displays the sample trend data, and figure (bottom) displays the correlation coefficients for the three signals. Note how correlation values change in areas where the blood pressure signals vary irrespective of each other, as around the 35th and 40th hour. Note also how correlation values change in areas where there is little variation in the signals though the blood pressure values seem to change in relation to each other, as after the 42nd hour.
Figure 3.11 (top) again shows the correlation coefficients, while figure (middle) displays the calculated CDI. Notice how the CDI is high during the regions of potential clots by the transducer, such as around the 35th and 40th hours. Figure (bottom) shows the regions of the underlying signal, from figure 3.10, that are above the minimum level of variability.
Figure 3.12 illustrates the final output of the correlation detector. Figure (top) again shows the CDI values, while figure (middle) displays the CDI values after those associated with regions of low variability have been removed. The dashed line represents threshold-variable. Note how in this case it is very high, and the user defined maximum constant value of var-lim-con is used instead.

3.6 Combined Detector

The output of each individual detector is saved and combined in the last stage of the algorithm. The regions labeled as artifactual from each detector are combined, as are regions labeled as uncertain by the variation detector and the correlation detector. This process is shown in figure 3.13. There is a degree of overlap among the outputs as more than one subsection may detect the same artifacts.
In the simplest case, the outputs from each subsection are combined using a simple OR calculation. With this setting, time locations are labeled as artifactual by the entire algorithm if they have been labeled as such by only one detector. Other logic functions, such as the AND function, can be used on different combinations of the outputs. This allows the user to perhaps increase thresholds and decrease the sensitivity of a single detector while increasing the sensitivity of the algorithm as a whole. These variations will be discussed further in the final chapter.

Figure 3.13 shows a block diagram of the aggregate artifact detector.
Chapter 4 - Benchmarking

4.1 The Gold Standard

The algorithm presented in the previous chapter attempts to address the challenge of artifact detection in physiological trend data from many different fronts. This chapter extends discussion of the algorithm beyond its motivation and theory and describes the process to develop a means by which to measure the algorithm's effectiveness - to develop a method to quantify not only the quality of the algorithm as a whole but also to fine tune and optimize the individual parameters of the algorithm.

Ideally, one would pass an infinite amount of physiological data through the algorithm. The output of this process - a noisy, uncertain, or clean tag for each section of data - would then be verified or invalidated by an expert knowledgeable in discerning artifacts. Even were the input data sets to be finite in size, studying the amount of data needed for adequate testing of the algorithm would prove to be very time consuming. A need thus arises for a collection of data already annotated such that regions of the data have been classified as clean or artifactual. These annotations can then be used as a "gold standard" in comparison with the algorithm's output.

The following will describe such a database of signals and then present results based upon use of this database.

4.2 ABP Alarm Database

Zong analyzed ABP alarms by designing a fuzzy logic approach to assess the underlying ABP signal quality. His method for discerning false alarms examined the ABP waveforms, the associated ECG waveforms, and the relationship between both sets of
signals. This led to a calculated "signal quality indicator" that was used to filter out supposed false alarms. In order to validate the results of his work, Zong created a library of patient data that contained both annotated alarms and notations on whether each alarm was true or false. The validity of these alarms was determined by manually examining each individual alarm as well as the underlying signal during the time the alarm was triggered. Using this data set as a benchmark, Zong was able to develop and test his algorithms.\(^2\) Zong’s work with false alarms can be extended to provide a database of signals with which to evaluate the effectiveness of additional artifact-detecting algorithms.

As stated before, Zong’s analysis was performed using waveform data. Certain modifications were thus needed to create a suitable benchmark set for use in this thesis. In place of ABP waveform data, recordings of HR, diastolic ABP, mean ABP, and systolic ABP were obtained from the patient set used in Zong's analysis. These recordings were measured at a frequency of .97656Hz. In order to be comparable to the MIMIC trend data of interest in this study, the recordings were processed using a low-pass FIR filter and then downsampled to reach a sampling rate of once-per-minute. This process, performed using MATLAB and its \texttt{resample} function, resembles the method used to calculate the actual trend data in the MIMIC database.

In all, data from 34 patients was used in creating the modified benchmarking database. This resulted in a set of 411 true alarms and 151 false alarms with which to compare the outputs from the algorithm. Figures 4.1 and 4.2 illustrate both a sample of trend data for a particular patient as well as the associated alarms for the data set.
Figure 4.1 shows a sample patient data from the benchmark database resampled for use by the algorithm. Figure (bottom) indicates the locations of alarms.
Figure 4.2 shows all alarms separated into true alarms and false alarms.

Potential drawbacks exist in using the system as described above. The first is the fact that alarms used in Zong's work, as are those in most clinical settings, are based upon waveform data. While the resampled trend data used here retains general details of the underlying signal, subtle and fast changes in value can be filtered out. Hence, the features of the underlying data that caused an alarm to trigger may not be evident on a trend level. Not only may the algorithm be unable to effectively evaluate alarms on this level, such waveform-based notations may not be applicable to this research. This problem, that of the benchmark data's suitability, is tested and discussed further in section 4.3. In addition, using Zong's data supposes that a false alarm is indicative of an
artifact existing in the underlying data. Though in general this was found to be true, Zong notes that this is not applicable for every false alarm. However, the incidences of false alarms associated with uncorrupted data are few enough that they are henceforth assumed to be inconsequential. As such, an occurrence of a false alarm is synonymous with an occurrence of corrupted data in the underlying signal. Similarly, a true alarm will signify clean data during the period of the alarm.

Comparison of the algorithm’s output with Zong’s database can occur only during periods when an alarm has been triggered. Though alarms vary in length, they seldom last for an extended period of time. Nevertheless, these are the only regions where notations concerning the underlying signal quality, whether the alarm was true or false, exist. Such notations will be used as the basis for determining the validity of the algorithm’s output concerning a processed signal. Though this severely limits the range over which to examine data, the large number of alarms present allows for a suitable collection of data from which to draw results.

4.3 Preliminary Examination
Initial analysis using Zong’s signal/alarm database consisted of examining characteristics of the signals during regions when alarms were triggered, a procedure independent of processing by the algorithm. This effort focused specifically on determining apparent differences in the underlying signals during true and false alarms. As stated previously, these regions are considered equivalent to regions of clean and corrupted data, respectively.
The intent of this preliminary study was to examine not only whether false alarms stem from a noticeable discrepancy among signals of interest, but also whether these can be quantified. The motivation behind this step was two-fold - test the benchmarking database as well as the theory behind the algorithm described in the previous chapter. Complimentary to testing the algorithm, this work also provided insight into possible values for the algorithm input parameters. Signal characteristics of interest were the same as those used by the proposed algorithm. Thus heart rate values, blood pressure values, heart rate variation, blood pressure variation, and blood pressure correlation were examined.

4.3.1 Missing Data Detection
The modified trend data version of Zong’s database exhibits no missing data as described in section 3.2. As was stated there, characteristics of missing data are highly dependent upon the mechanisms used to record the signals. The missing data subsection as described is designed specifically for use with the patient trend data stored in the MIMIC database. With the differences in measuring and sampling mechanisms used with Zong’s data, many of which were not fully known, it was impossible to determine analogous information pertinent to the MIMIC database. This detector was thus ignored at this point in time. This was allowable due to the fact that the missing data detector is the lone subsection of the algorithm with a known and highly defined input parameter set. Analysis of Zong’s data instead began by looking at characteristics used by the next subsection of the algorithm - the threshold detector.
4.3.2 Threshold Detection

Threshold detection assumes that certain artifacts are characterized by signal values that are markedly too high or too low. In reference to the benchmarking database, signal values would thus be excessive during certain false alarms. An examination of signal values during true and false alarms was performed in order to verify this assumption. The results are shown in Figures 4.3 and 4.4.

Figure 4.3 shows systolic blood pressure and mean blood pressure during alarms from Zong’s database.
Blood pressure values exhibit a discernible difference during true and false alarms, as exhibited in the above figures. Note how, though values during true alarms are clustered in a small area, values during false alarms extend into the upper right-hand corner of the graphs in both figures. This indicates that the signals during false alarms experience noticeably higher values. For example, whereas the mean blood pressure may reach a maximum value of slightly over 150 mmHg during true alarms, it can reach values greater than 350 mmHg during false alarms. A similar occurrence is true for both the systolic and diastolic blood pressures, as well. In each case, there is a noticeable difference.
4.3.3 Variation Detection

Variation detection assumes that certain artifacts are characterized by signal value transitions that are markedly too large. Again, in reference to Zong’s database, signal values would thus vary excessively during certain false alarms. A moving standard deviation was calculated over all blood pressure signals in order to examine variation events. Though many window-size values for the moving standard deviation were used in this step, all produced comparable results. As such, only variations over a window-size of three minutes will be discussed. The results of examining variations during true and false alarms are shown in Figures 4.5 and 4.6.

![Graph showing systolic and mean blood pressure variation](image)

*Figure 4.5 shows systolic and mean blood pressure variation (in terms of multiples of standard deviation above the mean) as calculated using a 3-minute window.*
Figure 4.6 shows systolic and diastolic blood pressure variation (in terms of multiples of standard deviation above the mean) as calculated using a 3-minute window.

Similar to the results of the threshold examination, the previous two figures indicate that blood pressure variation during true alarms tends to cluster around a limited set of values, whereas blood pressure variation during false alarms often reaches much higher values. This is indicated by the large number of values located in the upper right-hand corner of the graphs for false alarms. Once again using mean blood pressure as an example, standard deviation during true alarms tapers off after a value of 10. In contrast, the standard deviation of mean blood pressure during false alarms extends well past 50. Once more, there is a distinguishable difference between true and false alarms.
Discussion as to whether the signal differentiation between true and false alarms varies from that as established by the threshold examination, and whether there is value added by each detector, is saved for section 4.4.

4.3.4 Correlation Detection

Correlation detection assumes that many physiological signals are inherently correlated, and certain artifacts are characterized by signals that transition contrary to that as expected based upon the activity of other signals. Again, in reference to Zong’s database, particular signal values would thus vary completely irrespective of one another during certain false alarms. The moving correlations between systolic, mean, and diastolic blood pressures were calculated, and the corresponding CDI values for all alarms are shown in Figure 4.7.
Figure 4.7 shows systolic blood pressure variation in terms of standard deviation in contrast to the CDI values. Note the predominance of high CDI values clustered around low blood pressure variations during true alarms. This differs from the values during false alarms, where high CDI values exist among a broader range of variation values. Were the regions of low variability among blood pressure to be excluded from the CDI calculation, CDI values would present a better method of differentiating between true and false alarms among those noted in the benchmarking database.
4.3.5 Summary

The benchmarking database sufficiently displays differences in signal quality during true and false alarms, and thus during clean and corrupted regions of data. Note, however, that any single method of detection is unable to fully differentiate between true and false alarms. In each of the figures above, numerous false alarms populate the area typified by true alarms.

The CDI value, for example, is not high for all false alarms. A large proportion of false alarms have a CDI value close to zero, a trait similar to a predominant number of true alarms. However, each subsection of the algorithm is designed to detect different types of artifacts. In this situation, many locations with low CDI have high systolic variation, and these will most likely be detected during the variation detector. As such, the benchmark database has shown that it contains signals that are markedly different in regions of true and false alarms and was thus used in the next phase of this project as a standard from which to evaluate the algorithm.

4.4 Calibration of Algorithm Parameter Values

The availability of a suitable benchmark database allows for evaluation of the algorithm and its individual parameters. A method was thus developed to measure the algorithm’s ability to discern corrupted regions of data annotated in the benchmark database. Two values, sensitivity and positive predictivity, were used to quantify the algorithm’s performance.

Sensitivity, which measures the number of detected artifacts by the algorithm in relation to the overall number of artifacts, is used to evaluate the frequency with which
the algorithm is able to identify instances of corrupted data. It is defined, for use in this study, as follows.

\[
\text{Sensitivity} = \frac{\text{# of corrupted regions in the benchmark database tagged by the algorithm}}{\text{Total # of corrupted regions in the benchmark database}}
\]

4.1

Of greater importance than simple detection is the accuracy with which the algorithm can distinguish artifacts from true physiological events. It is possible for an algorithm to reach a perfect level of sensitivity by tagging every region of data as corrupted irregardless of its actual state of noise. This solution results in a large number of false negatives and is not feasible for use in areas such as an ICU. In such cases, a false negative is of much greater severity than a false positive.

Positive predictivity, which measures the number of correctly detected artifacts in relation to the overall number of tags, is used to evaluate the correctness of the algorithm output. It is defined, for use in this study, as follows.

\[
\text{Positive Predictivity} = \frac{\# \text{ of corrupted regions in the database tagged by the algorithm}}{\# \text{ of corrupted regions in the database tagged by the algorithm} + \# \text{ of uncorrupted regions in the database tagged by the algorithm}}
\]

4.2

The subsections of the algorithm were optimized in order to maximize these two attributes in combination. This process was simplified by optimizing each subsection independently. Data from the 34 patients in the benchmark database were passed through individual components of the algorithm as the detector parameters were varied incrementally. Sensitivity and positive predictivity were recorded for each batch run, and the results for each subsection are provided below.
4.4.1 Missing Data Detector

The missing data detector was omitted from the optimization process. As described in section 3.2, and alluded to again in section 4.3.1, its input is a single, non-varying parameter and thus not subject to optimization.

4.4.2 Threshold Detector

The threshold detector has three sets of parameters - limits for the systolic, mean, and diastolic blood pressures. As discussed in section 3.3, an understanding of the underlying signals is necessary when establishing such boundaries. A general review of the literature and discussion with experts helped determine 300, 250, and 200 mmHg as reasonable thresholds for use as the maximum limit for systolic, mean, and diastolic blood pressures, respectively. These were used as initial values during the optimization process. The results of calibrating the threshold detector using the benchmark database, with the optimal limits encircled, are shown in figures 4.8, 4.9, and 4.10.
Figure 4.8 shows the algorithm results while varying the threshold for systolic blood pressure. The threshold values increase from the right of the graph to the left by 10mmHg for each data point. Note how certain threshold values result in the same sensitivity and positive predictivity values. In the case of the systolic threshold parameter, sensitivity remains almost constant for all values above 300mmHg.
Figure 4.9 shows the algorithm results while varying the threshold for mean blood pressure. The threshold values increase from the right of the graph to the left by 10mmHg for each data point. Note again that certain values of the threshold, in this case 260, 270, and 280 mmHg for the mean maximum value, result in the same sensitivity and positive predictivity values.
Figure 4.10 shows the algorithm results while varying the threshold for diastolic blood pressure. The threshold values increase from the right of the graph to the left by 10mmHg for each data point.

As indicated above, the selected values for the systolic, mean, and diastolic limits were 300, 260, and 190mmHg, respectively. Though other thresholds may have provided a better combination of specificity and positive predictivity, achieving higher values for these statistics was of less importance than maintaining limits that were believed to be physiologically reasonable. Figure 4.10 shows that a diastolic maximum threshold of 130mmHg would produce both a higher sensitivity as well as a higher positive predictivity. However, as was determined at the beginning of this section, this value may be too low such that actual physiological data will be filtered out. As such, threshold values were allowed only limited flexibility to vary from the values determined prior to calibration. Thus, for values near the initial 200mmHg, 190mmHg provides the best...
outcome in terms of sensitivity and positive predictivity. Though a threshold of 180mmHg produces the same result, a value nearer to the original 200mmHg is preferred. Parameter values for systolic and mean maximum thresholds were determined using a similar methodology.

The threshold detector is the only detector for which there was less reliance on maximizing specificity and positive predictivity during calibration. With the aforementioned input parameters, it was able to detect 39 of the false alarms as annotated in the benchmark database. In addition, the detector mislabeled six true alarms as false.

4.4.3 Variation Detector

The variation detector is more difficult to optimize due to the interrelationships between its parameters. Whereas, for example, a maximum limit for diastolic threshold can be determined independently of the remaining threshold detector parameter values, the limit for variation is highly dependent upon the window size chosen for the moving standard deviation.

The window-size parameter was calibrated by examining its value over a large range of var_limmul. This analysis is illustrated in figures 4.11, and 4.12.
Figure 4.11 shows the algorithm results while varying the window size for the moving standard deviation calculation. Var_limmul increases from a value of 1 on the right of the graph to a value of 10 on the left. Note how, for all values of var_limmul, a window size of three is the optimum single value for window size.
Figure 4.12 shows the algorithm results while varying the window size for the moving standard deviation calculation. \texttt{Var\_limmul} increases from a value of 1 on the right of the graph to a value of 10 on the left. Note how an OR combination of a window size of three and a window size of five provides the optimal result. This combination is always more sensitive than with a window size of three, and it has similar positive predictive power for larger values of \texttt{var\_limmul}.

Figure 4.12 compares algorithm results from using multiple window sizes with that of using the optimum single window size of three minutes. Use of additional lengths increases the sensitivity without sacrificing an inordinate amount of positive predictive power. Thus, a combination of three- and five-minute windows was used. Parameter values for variation limits were calibrated after the optimal window size was determined. The results of \texttt{var\_limmul} optimization using the benchmark database, with the optimal limits encircled, are shown in figures 4.13, 4.14, and 4.15.
Figure 4.13 shows the algorithm results while varying var_limmul for systolic blood pressure. Var_limmul increases from a value of 1 on the right of the graph to a value of 10 on the left.
Figure 4.14 shows the algorithm results while varying \texttt{var\_limmul} for mean blood pressure. \texttt{Var\_limmul} increases from a value of 1 on the right of the graph to a value of 10 on the left.
A parameter value of 4 was chosen for the var_limmul value associated with each blood pressure signal. Unlike with the threshold detector parameters, variation detector parameters were chosen solely based on the resulting combination of sensitivity and positive predictivity. Var_limcon was similarly calibrated with respect to the selected window size. Figure 4.16, 4.17, and 4.18 show the results of the var_limcon optimization.
Figure 4.16 shows the algorithm results while varying varLimcon for systolic blood pressure.
Figure 4.17 shows the algorithm results while varying var_limcon for mean blood pressure.
Figure 4.18 shows the algorithm results while varying var_limcon for mean blood pressure.

The systolic var_limcon value was set to 21, mean to 27, and diastolic to 23. Again, these values produced the optimal combination of sensitivity and positive predictivity values. The variation detector, with these determined parameter values, was able to detect 90 of the false alarms as annotated in the benchmark database. In addition, the detector mislabeled 29 true alarms as false.

4.4.4 Correlation Detector

Similar to the variation detector, the correlation detector has parameters interrelated with each other. Limiting the value of variability in the underlying signal necessarily changes
the resulting range of CDI values. Figures 4.19 and 4.20 show the results of calibrating
the CDI value over a range of limits for blood pressure standard deviation.

![Correlation Detector Set Limit for CDI Calculation; Range of CDI=.1-2](image)

Figure 4.19 shows the algorithm results while varying the standard deviation limit for a given value of CDI. CDI increases from a value of .1 on the right of the graph to a value of 2 on the left in increments of .1 between each data point. Note how the positive predictivity increases as the minimum amount of standard deviation is raised.
Figure 4.20 shows the algorithm results while varying the standard deviation limit for a given value of CDI. CDI increases from a value of .1 on the right of the graph to a value of 2 on the left in increments of .1 between each data point. The optimum combination of sensitivity and positive predictivity is labeled.

Using the above graphs, values of 7.5 and .6 were chosen for the standard deviation minimum limit and CDI maximum limit, respectively, in order to best optimize sensitivity and positive predictivity. Using the determined value for standard deviation, the window size parameter was then calculated over a range of CDI values. The results of optimizing the window size value are shown in figure 4.21. As with the variation detector, results with the benchmark database showed that the use of multiple window sizes is optimal. As indicated by the encircled value, a combination of five-, ten-, and fifteen-minute window sizes results in the optimal set of positive predictivity and sensitivity.
Correlation Detector Set Limit for CDI Calculation; Range of CDI=0.1-2

Window Size=5
Window Size=10
Window Size=5 and 10
Window Size=5,10, and 15

Figure 4.21 shows the algorithm results with varying combinations of window sizes. CDI increases from a value of .1 on the right of the graph to a value of 2 on the left in increments of .1 between each data point.

The correlation detector, with the parameter values shown above, was able to detect 48 of the false alarms as annotated in the benchmark database. In addition, the detector mislabeled 32 true alarms as false.
4.4.5 Overall Results

Using the parameter values described above, the overall algorithm was able to correctly tag 118 instances of noise. However, 60 instances of true data were incorrectly tagged as artifactual. The results of the optimization process are summarized in table 4.1

<table>
<thead>
<tr>
<th>Detector</th>
<th># of Corrupted Regions Tagged by Algorithm</th>
<th># of Uncorrupted Regions Tagged by Algorithm</th>
<th>Sensitivity</th>
<th>Positive Predictivity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Threshold Detector</td>
<td>39</td>
<td>6</td>
<td>25.83%</td>
<td>86.67%</td>
</tr>
<tr>
<td>Variation Detector</td>
<td>90</td>
<td>29</td>
<td>59.60%</td>
<td>75.63%</td>
</tr>
<tr>
<td>Threshold and Variation Detector</td>
<td>92</td>
<td>30</td>
<td>60.93%</td>
<td>75.41%</td>
</tr>
<tr>
<td>Correlation Detector</td>
<td>48</td>
<td>32</td>
<td>31.79%</td>
<td>60.00%</td>
</tr>
<tr>
<td>Overall Combined Detector</td>
<td>118</td>
<td>60</td>
<td>78.15%</td>
<td>66.29%</td>
</tr>
</tbody>
</table>

Table 4.1 shows the results of optimizing the individual subsections of the algorithm using the benchmark database.
Chapter 5 - Results and Discussion

5.1 Summary of Results

The results of the algorithm calibration using the benchmark database are far from ideal. A sensitivity value of 78% indicates that almost a quarter of the false alarms are not detected. In addition, a positive predictivity of 66% is only slightly more accurate than randomly assigning a clean or corrupted label to any region of data.

Examining table 4.1, the threshold detector had the lowest sensitivity among the algorithm subsections, and it was not able to greatly increase the sensitivity associated with the variation detector when the two were used in combination. Nevertheless, the threshold detector is implemented to detect certain extreme and long lasting artifacts. The entirety of such noise, as those shown in figure 1.4, is only detected by the threshold detector.

The correlation detector had the lowest positive predictivity, and it greatly reduced that of the entire algorithm when added to the overall detector. However, much of the problem, as described in the next section, lies predominantly with the benchmark database.

5.2 Benchmarking Database

A sample patient data recording is shown in figure 5.1. The algorithm results from processing this data set are shown in figure 5.2.
Figure 5.1 shows a sample patient trend recording along with the alarm annotations as included in the benchmark database.
Figure 5.2 shows the results after passing the data through the algorithm. Note how, when comparing the False Alarm portion of figure 5.1 with the Tagged portion of figure 5.2, the algorithm appears to ably detect a large proportion of artifacts. Note that the additional regions tagged by the algorithm which are not noted as either a true or false alarm in figure 5.1 are disregarded in this study.

Though the algorithm output, as displayed in figure 5.2 (bottom), seems to relate well to the false alarms in the benchmark database, a closer look reveals that this is not entirely the case. The algorithm is unable to detect the first false alarm, and this region of data is magnified and shown in figure 5.3. Figure 5.4 shows the underlying blood pressure waveform from the same time period.
Figure 5.3 shows the first instance of a false alarm as indicated in the benchmark database as well as the algorithm output during that time. Note how the underlying trend data does not appear to indicate any artifact exists in that region.
Figure 5.4 shows the underlying blood pressure waveform data for the same time period as shown in figure 5.3.

The blood pressure recording shown in figure 5.3 contains two regions of artifacts. The patient monitor filters out the first occurrence, around minute 1350, and no alarm is sounded. The second occurrence, around minute 1355, is not filtered, and a false alarm is sounded. Note how the first artifact, seen clearly in figure 5.4, is of greater length than that of the second artifact. As a result, the first artifact is translated well to the trend data level while the second artifact is not evident in the trend data.

The algorithm is able to detect the first artifact but unable to detect the second. As the second artifact is the only one noted in the benchmark database, the algorithm is not credited for detecting the first artifact and penalized for not detecting the second artifact. A similar result occurs at minute 1926. Again, the artifact in the waveform data
is of such brevity that it does not appear in the trend data level. This is shown in figure 5.5. The result of these cases, where artifacts are not evident on the trend data level, is a decrease in the sensitivity of the algorithm.

![Blood Pressure Trend Data](image)

Figure 5.5 shows a false alarm noted in the benchmark database where the underlying artifact is not evident in the trend data.

Data from the same patient also illustrates areas where characteristics of the underlying waveform affect positive predictivity. Figure 5.6 shows a region where a true alarm is in close proximity to an artifact, and figure 5.7 shows the underlying blood
pressure waveform during this period. The artifact in the blood pressure data around minute 1630, as shown in figure 5.7, is of great length and thus appears on the trend level. However, due to its use of a window-size in excess of one minute, the algorithm labels regions immediately surrounding an artifact as also corrupted. The uncorrupted data immediately following the artifact is thus included in this labeled region. Figure 5.8 illustrates an additional situation where the proximity of a true alarm to an artifact causes it to be tagged as noisy.
Figure 5.6 shows a true alarm noted in the benchmark database tagged as artifactual by the algorithm.
Figure 5.7 shows the underlying blood pressure waveform during the time period shown in figure 5.6
The previous figures fail to illustrate an additional drawback to using the benchmark database – the prior probabilities of true and false alarms are disproportionate to each other. The number of true alarms is greater than the number of false alarms by almost three-fold. Any calculation of positive predictivity may thus be biased towards the misinterpretation of true alarms. This may also help explain, in some part, the low value associated with the algorithm.

Unfortunately, though Zong’s database is well-suited for analysis of waveform data, an artifact detection algorithm using both the converted trend data signal set and the
associated waveform-based set of annotations ultimately proved to be inadequate for this analysis.

5.5 Conclusions and Future Work

Modern-day artifact detection is limited in its ability to detect noise in physiological parameter data. Use of higher level statistical analysis has been shown to complement current rule based filters in analyzing variations among physiological signals. More importantly, however, is the incorporation of relationships between signals into the algorithm. Such inter-signal correlations are effective in detecting additional regions of noise in the trend data. However, additional work is needed to further determine the true effect of this study.

There is a need to explore additional annotated physiological signal databases for use in calibration of the algorithm. Such a database would necessarily include artifact notations applicable specifically on a trend level. Other physiological signals may also provide additional insight into methods for calibrating specific detector parameters. Further work may thus include the analysis of physiological signals such as CVP and DPAP.

The algorithm described in this thesis may also be extended in many ways. Potential research may include alternative methods and logic functions for combining the tagged output data sets from each subsections as well as different methods for the handling of missing data.

Finally, work on this thesis has provided methods that may be useful in detecting specific events. One such event, that of a clot in the catheter, is often difficult to observe.
However, its effects are readily apparent on a trend level. Use of a modified version of the correlation detector may be used to develop an algorithm specifically suited for clot detection. Artifact detection is of vital importance to both the clinical and research environment. It is hoped that this study will be able to help advance the use of physiological data in both settings.
Bibliography


15 http://www.aboutatrialfibrillation.com/whatisaf.html