



BOSCH
Technik fürs Leben



Duale Hochschule
Baden-Württemberg

Entwicklung eines Tools zur Auswertung medizinischer Messdaten für eine Pilotstudie im Bereich der Stress-Regulationsfähigkeit des Menschen

Studienarbeit

Bachelor of Science

des Studiengangs Informatik

an der Dualen Hochschule Baden-Württemberg Stuttgart

von

Jona Krumrein, Tim Weiss

10.06.2022

Bearbeitungszeitraum
Matrikelnummer, Kurs
Ausbildungsfirma
Betreuer

22.10.2022 - 10.06.2022
3857223, 8336074, TINF19ITA
Robert Bosch GmbH, Stuttgart
Mario Babilon

Abstract

TODO: deutscher Abstract....

Inhaltsverzeichnis

Abkürzungsverzeichnis	IV
Abbildungsverzeichnis	V
Tabellenverzeichnis	VI
Formelgrößenverzeichnis	VII
Formelverzeichnis	VIII
Listings	IX
1 Medizinische Grundlagen	1
1.1 Herzparameter	1
1.2 Herzfrequenzvariabilität (HRV)	3
2 Aufgabenstellung	4
3 Stand der Technik	5
3.1 Kubios HRV Premium	5
3.2 Vergleich möglicher Datenstrukturen der Messdaten	5
3.3 Auswertung von Messdaten mit MATLAB und Python	6
3.3.1 MATLAB vs. Python	7
3.3.2 Performance-Vergleich	7
4 Konzept	10
5 Implementierung des Tools	11
5.1 Grundstruktur der Applikation	11
5.2 Erstellung der Samples	11
5.3 Visualisierung der Daten	11
5.3.1 Datenstruktur	11
5.3.2 Plot-Funktion	11
5.3.3 Zusätzliche Funktionen	11
Anhang	A

Glossar

B

Abkürzungsverzeichnis

BSP Board Support Package

Abbildungsverzeichnis

1.1 R-R-Intervall	3
-----------------------------	---

Tabellenverzeichnis

3.1	Übersicht der technischen Daten des Testmediums	7
3.2	Technische Daten	8

Formelgrößenverzeichnis

a	rad	Bedeutung von a
b	rad	Bedeutung von b
λ	rad	Bedeutung von lambda
ϕ	rad	Bedeutung von phi

Formelverzeichnis

1.1 Herzminutenvolumen (HMV)	1
1.2 Maximale Herzfrequenz	1

Listings

3.1	MATLAB Implementierung	8
3.2	Python Implementierung	8

1 Medizinische Grundlagen

1.1 Herzparameter

Das menschliche Herz kann mit vielen verschiedenen Werten parametrisiert werden. Der wohl bekannteste Parameter ist dabei die Herzfrequenz (H_f). Sie gibt an, wie oft das Herz in einer bestimmten Zeit schlägt. Meist wird dafür das Zeitintervall einer Minute gewählt und die in dieser Zeit gemessenen Herzschläge in der Einheit *bpm* (beats per minute) angegeben.

Die Herzfrequenz ist abhängig von verschiedenen Faktoren. So spielen beispielsweise Alter und körperliche Fitness eine große Rolle. Bei trainierten (Ausdauer) Sportlern schlägt das Herz in Ruhe bedeutend seltener als bei untrainierten Personen. Dies kann damit erklärt werden, dass durch das Ausdauertraining das **Schlagvolumen (SV)** erhöht werden kann. Das Schlagvolumen gibt an, wie viel Blut das Herz mit einem Schlag durch den Körper pumpen kann. Wenn sich dieses Volumen also erhöht, muss das Herz seltener schlagen, um das **Herzminutenvolumen (HMV)** zu halten. Dieses berechnet sich aus Herzfrequenz und Schlagvolumen:

$$H MV = H f * S V \quad (1.1)$$

Bei körperlicher Belastung, beispielsweise sportlicher Aktivität, erhöht sich das benötigte Herzminutenvolumen, da die Muskeln mehr Sauerstoff brauchen. Da sich das Schlagvolumen des Herzens nicht anpassen lässt, erhöht sich die Herzfrequenz um dem benötigten Herzminutenvolumen gerecht zu werden. Grundsätzlich, wie jeder bei sich selber feststellen kann, erhöht sich die Herzfrequenz bei körperlicher Aktivität. Durch sehr starker Aktivität kann die Herzfrequenz auf ein Maximum ansteigen. Die maximale Herzfrequenz kann mit folgender Faustformel abgeschätzt werden:

$$H f_{max} = 220 - A l t e r \quad (1.2)$$

gesprochen. Allerdings muss zwischen Puls und Herzfrequenz unterschieden werden, da diese nicht identisch sind. Die Herzfrequenz beschreibt, wie bereits in den vorhergehenden

Abschnitten beschrieben, die tatsächlichen Schläge des Herzens. Zur Messung der Herzfrequenz kann ein EKG genutzt werden, Der **Puls** hingegen wird peripher, beispielsweise am Hals oder Handgelenk gemessen. Dabei werden Pulswellen erfasst, welche sich durch den Körper bewegen. Eine Pulswelle entsteht wenn das Herz schlägt, und das Blut durch den Körper pumpt. Dabei wird das Blut gegen die Arterienwände gedrückt und eine Pulswelle kann detektiert werden. Bei gesunden Menschen entspricht der Puls der Herzfrequenz, da jeder Herzschlag Blut durch den Körper pumpt und somit auch eine messbare Pulswelle erzeugt. Bestimmte Herzrhythmuskrankheiten können dafür verantwortlich sein, dass Puls und Herzfrequenz voneinander abweichen. Dabei können Kontraktionen des Herzens entstehen, welche kein oder nicht genug Blut durch den Körper pumpen um eine messbare Pulswelle zu erzeugen. Die Differenz zwischen Herzfrequenz und Puls nennt sich **Pulsdefizit** (und ist bei gesunden Menschen 0).

Die Variation der Herzfrequenz wird über das **vegetative Nervensystem** geregelt. Dies besteht aus dem *Sympathikus* und seinem Gegenspieler, dem *Parasympathikus*. Außerdem ist das *enterische Nervensystem* (*Eingeweidenervensystem*) teil des vegetativen Nervensystems. Dies spielt für die Steuerung des Herzens allerdings keine Rolle und wird daher im Laufe dieser Arbeit nicht mehr erwähnt. Ebenso haben Sympathikus und Parasympathikus Funktionen in der Verdauung. Da dies ebenfalls keine Relevanz für das Herz hat, wird darauf nicht weiter eingegangen. Die Folgende Erklärung von Sympathikus und Parasympathikus beziehen sich daher nur auf die für das Herz relevanten Eigenschaften.

1. Sympathikus

Der Sympathikus kann den Organismus auf eine Aktivitätssteigerung ("fight or flight") einstellen. Er ist der dominante Teil des vegetativen Nervensystems, wenn es um Stresssituationen geht. Wenn diese eintritt, wird die Aktivität von situationsbedingt unwichtigen Organen, beispielsweise Darm oder andere Organe die zur Verdauung beitragen, gesenkt. Die Aktivität von Organen die in Stresssituationen wichtig sind wird verstärkt. So werden beispielsweise die Pupillen vergrößert. **Die Herzfrequenz wird stark erhöht.**

2. Parasympathikus

Der Parasympathikus stellt den Organismus auf Ruhesituationen ("rest and digest") ein. So werden in den dominanten Phasen des Parasympathikus Organe mit Verdauungsfunktionen in ihrer Aktivität gestärkt. Der Körper wird auf Entspannung und Regeneration heruntergefahren. **Die Herzfrequenz wird in diesen Phasen stark gesenkt.**

Das vegetative Nervensystem ist in seiner Funktion **unwillkürlich**. Das bedeutet, es kann nicht bewusst von der jeweiligen Person gesteuert werden. Dadurch sind Messungen, beispielsweise an der Herzfrequenz sehr gut geeignet, um bestimmte äußere Einflüsse zu untersuchen. Die Ergebnisse können nicht von der Testperson verfälscht werden, da sie das vegetative Nervensystem nicht beeinflussen kann. (Andere äußere Einflüsse wie Aufregung müssen berücksichtigt werden und können das Ergebnis logischerweise verfälschen.)

Um nun den konkreten Einfluss von elektromagnetischen Wellen zu untersuchen kann die Herzfrequenz als Messparameter benutzt werden. Ein weiterer Parameter welcher stark abhängig vom vegetativen Nervensystem ist, ist die **Herzfrequenzvariabilität (HRV)**. Im Verlauf dieser Arbeit, wird die Herzratenvariabilität als entscheidender Parameter genutzt. Alle durchgeführten Messungen und daraus abgeleiteten Interpretationen beziehen sich auf diese.

1.2 Herzfrequenzvariabilität (HRV)

Die Herzfrequenzvariabilität, auch Herzratenvariabilität ist die zeitliche Variation zwischen zwei aufeinanderfolgenden Herzschlägen [wiki?](#).

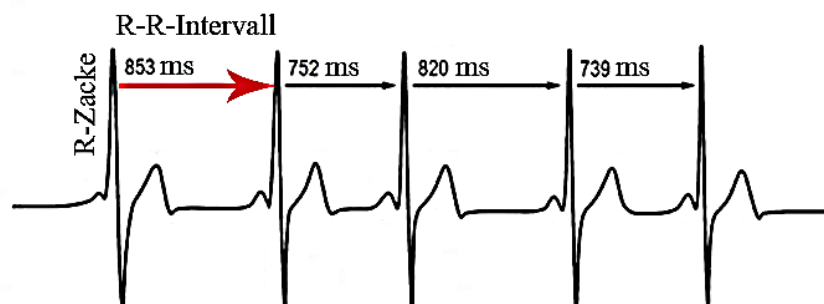


Abbildung 1.1: R-R-Intervall

In dieser Grafik ist eine EKG Messung zu sehen. Gut erkennbar sind die hohen R Zacken. Diese werden benutzt, um die Herzfrequenzvariabilität zu berechnen. Dabei wird die Zeit zwischen zwei Herzschlägen als die Zeit zwischen den aufeinanderfolgenden R-Zacken definiert. In diesem Schaubild liegt diese Zeit, also das R(-)R-Intervall zwischen 739 ms und 853 ms. Das RR-Intervall wird in der Literatur oft als NN-Intervall bezeichnet. Der Begriff RR wird auch bei Messen des Blutdrucks verwendet. Dort steht RR für *Riva-Rocci*. Um eine Verwechslung zu vermeiden, wird daher im Kontext der HRV vom NN-Intervall gesprochen.

2 Aufgabenstellung

Mit der fortschreitenden Entwicklung in der Digitalisierung unserer Gesellschaft erhöht sich ebenfalls die Anzahl und Verfügbarkeit der notwendigen Geräte und der dazugehörigen Sendeanlagen. Diese können mit Hilfe von elektromagnetischen Wellen kabellos Daten übertragen und erzeugen somit künstlich elektromagnetische Felder, welche sich mit den bereits natürlichen elektromagnetischen Feldern überlagern. Die Folgen des wachsenden Ausbaus digitaler kabelloser Technologien, sowie die Nutzung der passenden Endgeräte, wie Smartphones, Tablets und drahtlose Telefone und die damit verbundenen Auswirkung auf die Bevölkerung, rückte bisher nur in geringem Maße in den Fokus der wissenschaftlichen Forschung.

Innerhalb einer Pilotstudie sollen die Auswirkungen und Einflüsse elektromagnetischer Strahlung auf die Stress- und Reizverarbeitung des Menschen untersucht werden. Dazu werden verschiedene biologische Messwerte des Herzens ohne Strahlenbelastung bestimmt und danach den alltäglichen Belastungen durch WLAN-Router, DECT-Telefone, Bluetooth- und Mobilfunk-Endgeräte gegenübergestellt. Bei der Durchführung der verschiedenen Messungen und Untersuchungsarten der Studie wird eine große Datenmenge produziert, welche passend aufbereitet werden muss, um eine schnelle und einfache Auswertung durch die beteiligten Ärzte und Physiotherapeuten ermöglichen zu können.

Ziel der Studienarbeit ist die Konzeptionierung und Implementierung einer Schnittstelle, welche die Automatisierung der Aufbereitung und Auswertung der Messdaten ermöglicht. Dazu soll eine Applikation erstellt werden, welche an die bereits verwendete und etablierte Software Kubios HRV Premium anknüpft und deren Funktionen erweitert. So sollen die Rohdaten der Messungen in Kubios eingelesen und die Exportdateien und deren Herzparameter innerhalb einer Applikation graphisch dargestellt werden. Der Fokus liegt dabei auf der automatische Generierung der in Kubios erstellbaren Samples, welche die Einteilung der Messung in kleinere Zeitbereiche ermöglicht, sowie der fachlich korrekten Darstellung der Herzparameter im Diagramm zur Unterstützung der Studie. Zudem soll die Möglichkeit bestehen das Belastungsintervall hervorzuheben und die fertige Visualisierung zu speichern. Zur Umsetzung sollen dabei ebenfalls unterschiedliche Programmiersprachen sowie Dateiformate betrachtet und bewertet werden.

3 Stand der Technik

3.1 Kubios HRV Premium

3.2 Vergleich möglicher Datenstrukturen der Messdaten

Wie bereits im Kapitel zuvor erwähnt bietet Kubios HRV Premium zwei unterschiedliche Exportmöglichkeiten. Diese werden durch den einheitlichen CSV-Standard und das MATLAB-eigenen MAT-File repräsentiert.

Bei einer CSV-Datei (engl. Comma-separated values) handelt es sich um einen Standard, welcher den Aufbau einer Textdatei beschreibt. Der Inhalt der Datei beschränkt sich meist auf Tabellen und Listen, welche nicht wie üblicherweise in Zeilen und Spalten aufgeteilt wird, sondern mit Hilfe von Kommas getrennt wird, wobei jede neue Zeile eine neue Datenbankzeile darstellt und jede Datenbankzeile aus einem oder mehreren Feldern besteht. CSV-Dateien finden vor allem Anwendung in Tabellenkalkulationsprogrammen und Datenbanksysteme, wie Microsoft Excel oder MySQL, welche das Format einlesen, sowie exportieren können. Aufgrund der weiten Verbreitung des Formats ist auch das Einlesen und Verarbeiten der Dateien in nahezu jeglicher Programmiersprache möglich. Vor allem Python bietet passende Bibliotheken und Funktionen.

Das MAT-File beschreibt ein binäres Datencontainerformat, welches von MATLAB verwendet wird. Es ermöglicht das Speichern von Arrays, Variablen und Funktionen. Je nach MATLAB Version ermöglicht das Format zusätzlich die Speicherung von mehrdimensionale numerische Arrays, Zeichenarrays, Zellenarrays, Felder mit geringer Dichte, Objekte und Strukturen. Die Lesbarkeit beschränkt sich hier jedoch auch auf MATLAB und einige wenige Bibliotheken für beispielsweise Python.

Da diese beiden Export-Formate und deren Inhalt von einander verschieden sind und die Grundlage der Datenverarbeitung sowie Datenvisualisierung bilden, sollen diese im Folgenden Hinblick auf Übersichtlichkeit, Formatierung, inhaltlicher Korrektheit und Zugänglichkeit verglichen werden.

Die Darstellung der CSV-Datei ist unverschachtelt und befindet sich auf einer Ebene. Alle Daten lassen sich auf einmal einsehen. Innerhalb der Datei befinden sich tabellarische Strukturen der einzelnen Messdaten, aber auch Einzelinformationen, Überschriften und Bezeichner. Dies stellt für das Einlesen und Verarbeiten der Daten ein Problem dar, da zusätzlich nach relevanten gefiltert oder mehrere Formatierungsiterationen durchgeführt werden müssen. Das MAT-File hingegen ist in sich verschachtelt. Die Messdaten sind sortiert und unter ihrem jeweiligen Bereich geordnet. Der Zugriff auf die Daten erfolgt mittels Key-Value-Verfahren, das heißt unter einem Schlüsselwort befinden sich der jeweilige Datensatz. Ein Datensatz kann dabei ebenfalls von einem weiteren Key-Value-Paar dargestellt werden. Irrelevante Daten können hier gefiltert werden in dem man den jeweiligen Key einfach nicht aufruft. Das Format kann also ohne weitere Vorformatierung eingelesen werden.

Betrachtet man die Konstanz und Korrektheit bezogen auf die Darstellung der Daten in Kubios, lassen sich im CSV-Format einige Unregelmäßigkeiten erkennen. So sind häufig Kommas verschoben und Felder mit einer stark variierenden Anzahl Leerzeichen aufgefüllt, um eine übersichtlichere Darstellung zu ermöglichen. Dies hat jedoch den Effekt, dass diese Leerzeichen beim Einlesen mit beispielsweise Python ebenfalls im Datensatz auftauchen und diesen verunreinigen bzw. im schlimmsten Fall unbrauchbar machen. Auch hier müsste deshalb im Vorhinein eine Formatierung der grundlegenden Datei durchgeführt werden. **Diese Unregelmäßigkeiten lassen sich auf die Exportfunktionen der Kubios-Software zurückführen.** Der Inhalt des MAT-Files ist hingegen akkurat und nicht von den Unregelmäßigkeiten betroffen.

Daraus ergibt sich, dass das MAT-File ein grundlegend konsistenterer Datenspeicher für die medizinischen Messdaten darstellt und einen großen Vorteil im Bereich der Zugänglichkeit sowie der Reduzierung zusätzlicher Hilfsfunktionen und Datenmanipulationen gegenüber dem CSV-Format bietet. Aus diesem Grund wird dieses auch als Grundlage zur Umsetzung des Tools zur Auswertung der medizinischen Messdaten verwendet.

3.3 Auswertung von Messdaten mit MATLAB und Python

Im Bereich der Data Science gibt es viele unterschiedliche Programmiersprachen, welche das Bearbeiten und Auswerten großer Datenmengen ermöglichen. Je nach Anwendungsgebiet und Lösungsansätzen, sowie der dazugehörigen Eigenschaften der zu analysierenden Daten muss eine passende Programmiersprache gewählt werden. Im Hinblick auf die Anforderungen der Studienarbeit stehen vor allem das Auslesen der Daten aus vorgegebenen

Datei-Formaten, das graphische Darstellen der Messdaten, sowie die Erstellung einer passenden Benutzeroberfläche im Vordergrund.

3.3.1 MATLAB vs. Python

3.3.2 Performance-Vergleich

Um einen besseren Überblick über die Performance der beiden Programmiersprachen zur Auswertung von Messdaten zu erhalten, werden mehrere Speedtests durchgeführt. Im ersten Schritt wird dazu ein kleiner Datensatz geöffnet und ein einfache Line-Plot erstellt. Geplottet wird jeweils die durchschnittliche Herzfrequenz aus dem normalen Datensatz der Auswertung ohne Samples. Hierbei handelt es sich um 41 Werte, welche als Line-Plot ohne weitere Konfiguration dargestellt werden. In der Python Umgebung werden die Bibliotheken **pandas** und **h5py** verwendet, da diese für ihre jeweiligen Aufgaben, Erstellen von Data Frames und Laden von mat-Files, als State-Of-The-Art gelten. Zum Plotten wird auf die **matplotlib** zurückgegriffen, da diese der MATLAB Darstellung am nächsten kommt. MATLAB ermöglicht die Auswertung ohne zusätzliche Bibliotheken. Die Zeit wird in beiden Fällen mit nativen Funktionen ausgewertet. Getestet wird auf einem MacBook mit folgenden technischen Daten. Dabei werden beide Skripts jeweils dreimal abwechselnd hintereinander ausgeführt, während sich das MacBook im Akku-betriebenen Modus befindet.

Technische Daten

Model	MacBook Pro (Retina, 13-inch, Early 2015)
Betriebssystem	macOS Monterey Version 12.1
Prozessor	2,7 GHz Dual-Core Intel Core i5
Arbeitsspeicher	8 GB 1867 MHz DD3
Grafikchip	Intel Iris Graphics 6100 1536 MB

Tabelle 3.1: Übersicht der technischen Daten des Testmediums

Die Implementierungen der einzelnen Programmiersprachen, sowie die berechneten Programmlaufzeiten sind im Folgenden dargestellt. Die beiden Skripte werden jeweils fünf mal ausgeführt und der Mittelwert der berechneten Zeiten bestimmt.

```
1 tic;
2 load(' ../dat/11-48-21_hrv.mat ');
3 plot(Res.HRV.TimeVar.mean_HR);
4 tac;
```

Listing 3.1: MATLAB Implementierung

```
1 import time
2 import pandas as pd
3 import h5py
4 import matplotlib.pyplot as plt
5
6 start = time.time()
7 f = h5py.File(' ../dat/11-48-21_hrv.mat ')
8 df = pd.DataFrame(f.get(' Res/HRV/TimeVar/mean_HR ')).T
9
10 df.plot(y=0, kind='line')
11 end = time.time()
12 print(end - start)
13 plt.show()
```

Listing 3.2: Python Implementierung

Tabelle 3.2: Technische Daten

Durchlauf	MATLAB	Python
1	0.303867 s	0.359526 s
2	0.307696 s	0.351593 s
3	0.295474 s	0.354158 s
4	0.301647 s	0.352104 s
5	0.299218 s	0.350182 s
Mittelwert	0.301580 s	0.353512 s

MATLAB ist bei jeder Ausführung um ca. 17% schneller als das Python Skript. Zudem muss die Komplexität der beiden Skripte betrachtet werden. In MATLAB benötigt man lediglich zwei Zeilen Code und keine zusätzlichen Bibliotheken, während das Python Skript vier Code Zeilen und drei zusätzlichen Bibliotheken in Anspruch nimmt. Außerdem

muss hier beachtet werden, dass das Anzeigen des Plots unter Python nicht mit in die Berechnung der Zeit aufgenommen werden kann, da alle Code-Zeilen nach „plt.show“ auch erst nach dem Schließen des Plot-Fensters angezeigt werden.

4 Konzept

5 Implementierung des Tools

5.1 Grundstruktur der Applikation

5.2 Erstellung der Samples

5.3 Visualisierung der Daten

5.3.1 Datenstruktur

5.3.2 Plot-Funktion

5.3.3 Zusätzliche Funktionen

Anhang

Glossar

Glossareintrag

Ein Glossar beschreibt verschiedenste Dinge in kurzen Worten.