

IDP im Sommersemester 2013

Quality Control Tool: Adaption to raw data

Sarah Barton und Sonja Vogl
Technische Universität München

05. Juli 2013

Zusammenfassung

Diese Ausarbeitung gibt dem Benutzer einen Einblick in die Bedienbarkeit des QualityControlTools. Ausgehend von der ersten Version der beiden Entwickler Martin Attenkofer und Johannes Eder werden die Erweiterungen des QualityControlTools vorgestellt. Zudem geht diese Arbeit auf die Funktionen ein, die das QualityControlTool zur Analyse von Daten der Biosensoren GeneActive, Somnowatch, Shimmer und GT3X+ bereitstellt.

Inhaltsverzeichnis

1	Einleitung	4
2	Motivation	4
3	Projektbeschreibung	4
4	Ausgangslage - die erste Version des Tools	5
5	Das erweiterte Tool	8
5.1	Der Browse-Button	9
5.1.1	Aktion des Benutzers	9
5.1.2	Aktionen des QualityControlTool	9
5.1.3	Implementierung	10
5.2	Der Load-Button	11
5.2.1	Möglichkeiten für den Benutzer	11
5.2.2	Implementierung	11
5.3	PlotAllData + Cut -Button	12
5.3.1	Implementierung	13
5.3.2	Cut Data	14
5.4	PlotRelevantData-Button	14
5.4.1	Implementierung	14
5.5	Wearing time van Hees-Button	15
5.5.1	Implementierung	15
5.6	Raw to Counts/METs-Button	16
5.6.1	Aktionen des QualityControlTool	16
5.6.2	Implementierung	16
5.7	Analyse nach Andre	17
5.7.1	Beschreibung	17
5.7.2	Implementierung	18
5.8	Analyse nach Simon	18
6	Abhängigkeiten im QualityControlTool	19
7	Beispielablauf einer Analyse von Shimmer Daten	20
7.1	Starten des Programms	20
7.2	Laden einer Shimmer Datei	21
7.3	Umwandeln von Rohdaten in Counts	23
7.4	Plotten aller Daten	24
7.5	Plotten der relevanten Rohdaten	26
7.6	Berechnung der Tragezeit nach van Hees	28
7.7	Epochenänderung	29
7.8	Analyse nach Andre	30
7.9	Analyse nach Simon	30
8	Hinweise zur Benutzung des Tools	30

9	Mögliche Verbesserungen und Erweiterungen	31
10	Fazit	32

1 Einleitung

Diese Arbeit gibt dem Benutzer des QualityControlTools einen Überblick über die Möglichkeiten, die das Programm zur Analyse der Biosensoren GeneActive, Somnowatch, Shimmer und GT3X+ bietet. Zudem wird dem Benutzer nahegebracht, wie das QualityControlTool zu bedienen ist. Hierzu wird zunächst die Benutzeroberfläche (siehe Kapitel 5), wie auch die Funktionsweise der einzelnen Buttons dargestellt. Zusätzlich wird grob auf die Implementierung eingegangen. Kapitel 6 bietet dann einen Überblick über die Zusammenhänge zwischen den einzelnen Buttons.

Um den Einstieg für neue Nutzer des Tools möglichst einfach zu gestalten wird in Kapitel 7 beschrieben wie eine Auswertung einer Datei ablaufen kann. Außerdem werden in Kapitel 8 einige nützliche Hinweise zur richtigen Benutzung des Tools gegeben.

2 Motivation

Das Institut für Medizinische Statistik und Epidemiologie (IMSE) der TU München beschäftigt sich mit der Analyse der Daten der KORA Studien. KORA steht für die *kooperative Gesundheitsforschung in der Region Augsburg*.

Das Ziel der *Feasibility*-Studie der Nationalen Kohorten war das Testen der Realisierbarkeit mono- und multifunktionaler Apparaturen, die Körperbewegungen messen. So sollten Körperleistung (PA) und der Energieverbrauch analysiert werden. Im speziellen wurden vier verschiedene Biosensoren betrachtet, die *raw acceleration data* für mindestens 7 Tage lieferten.

Das QualityControlTool soll das Auswerten der, von den Biosensoren erstellten, csv-Dateien erleichtern. Für die Auswertung existieren zahlreiche Algorithmen, welche in MATLAB implementiert sind. Allerdings besitzen viele Anwender nur ein eingeschränktes Informatikwissen. Für diese ist eine Auswertung mit einem hohen zeitlichen und hohem Arbeitsaufwand, sich in Matlab einzuarbeiten, verbunden.

Mit dem QualityControlTool ist eine einfache Bedienbarkeit ohne informatische bzw. mathematische Kenntnisse gegeben. Sobald man mit der Benutzeroberfläche vertraut ist, übernimmt der PC das Auswerten der Daten. Zusätzlich werden die ausgewerteten Daten so gespeichert, dass sie schnell auffindbar sind und jederzeit erneut abgerufen werden können.

3 Projektbeschreibung

Das bestehende QualityControlTool soll im Zuge dieser Arbeit für die Sensoren ActiGraph GT3X+, GeneActive, Somnowatch und Shimmer erweitert werden. Da Dateien dieser Sensoren nur in Rohdaten vorliegen, müssen die Funktionen

für diesen Datentyp erweitert werden. Dabei sollen alle Funktionen flexibel für die an dem Sensor eingestellten Frequenzen funktionieren.

Damit die Bedienbarkeit für den Anwender einfach bleibt soll außerdem die Oberfläche angepasst werden. Dazu wird die Auswahlleiste für den Sensortyp, sowie für den Datentyp aktiviert. Zudem soll dem User eine Möglichkeit zur Einstellung und zum Ablesen der Frequenz geboten werden.

Außerdem sollen die beiden neuen Funktionen *wearing time van Hees* und *Raw to Counts/METs* in das Tool integriert werden. Die Erste dient der Möglichkeit die Tragezeit von Rohdaten zu ermitteln. Die zweite Funktion soll die geladenen Rohdaten in Counts umwandeln, so dass mit diesen Daten genauso wie in der ersten Version des Tool gearbeitet werden können.

4 Ausgangslage - die erste Version des Tools

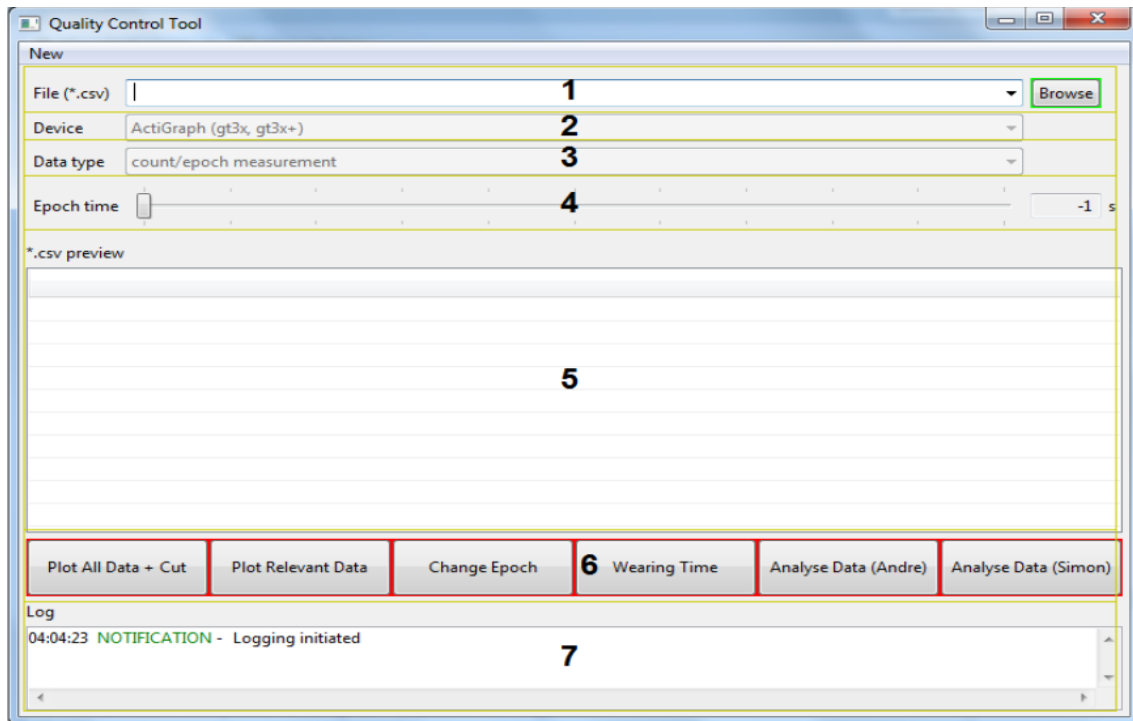


Abbildung 1: GUI des QualityControlTools 1. Version

Die erste Version des QualityControlTools¹ war bislang nur auf den Biosensor ActiGraph GT3X ausgerichtet. Für diesen Sensor ließen sich csv-Dateien des Datentyps *count data* auswerten. Hierzu konnte der Benutzer einen Pfad mit dem entsprechenden Dateinamen übergeben. Beim Betätigen des *Browse*-Buttons (siehe Abb. 1: 1) wurden die Daten der gewählten Datei geladen. Diese wurden dann in TIMESERIES-Format als MATLAB-Datei unter dem Namen *Kora1.mat* gespeichert.

Für den Sensortyp (*Device*) und den Datentyp (*Data type*) war bereits die Vorrichtung für eine Auswahlmöglichkeit des Benutzers vorhanden (siehe Abb. 1: 2 und 3). Diese blieb aber in der ersten Version dem Benutzer gesperrt. Das heißt die Anzeige war automatisch auf *Actigraph(gt3x,gt3x+)* sowie auf *count/epoch measurement* gesetzt und konnte nicht verändert werden.

Zudem ermöglichte es das Tool durch das Betätigen des *Plot All Data + Cut*-Button, aus den eingelesenen Daten eine Grafik zu erstellen. Der, aus diesem Plot, resultierende Graph wurde unter dem Namen *KoraFirstPlot.png* im Dateisystem gespeichert.

In einem nächsten Schritt konnten die gegebenen Daten zugeschnitten werden. Der Benutzer trug hierzu einen Starttag und die Anzahl der Folgetage in ein vorgegebenes Feld ein. Alle Daten, die sich nicht in diesem Zeitraum befanden, wurden weggesehnt. Die resultierende TIMESERIES überschrieb *Kora1.mat*. Aus den entstandenen Daten wurde wiederum ein Graph erstellt. Dieser wurde unter dem Namen *KoraSecondPlot.png* im Dateisystem abgespeichert.

Als ein so genannter Zwischenschritt war das Verändern der Epochenlänge zu sehen. Dafür musste zunächst die gewünschte Epochenlänge, als ein Vielfaches der bisherigen, eingestellt werden. (siehe Abb. 1: 4) Mit Hilfe von Matrizenberechnungen wurden die Werte eines Intervalls aufsummiert. Das Resultat überschrieb ebenfalls *Kora1.mat*.

Die Daten der csv-Datei ließen sich sowohl mit Hilfe von dem Algorithmus von Andre wie auch nach dem Algorithmus von Simon analysieren. Bei dem erstgenannten wurden für jeden Tag der Eingabedaten eine bestimmte Anzahl an Werte berechnet, welche in einem Vektor gespeichert wurden. Der Vektor ließ sich anschließend in der Datei *finaldata.csv* auffinden. Bei dem Algorithmus nach Simon wurden sowohl pro Tag, als auch für den gesamten Datensatz eine bestimmte Anzahl an Daten bestimmt. Die Ergebnisse waren in den Dateien *sum.table_[Klassenname]* und *sum.filt_[Klassenname]* abgelegt. Um Probleme in Java vorzubeugen, wurde am Ende der Workspace geleert.

Um jedoch die Analyse nach Simon berechnen zu können musste zuerst die

¹Johannes Eder, Martin Attenkofer, *IDP: Quality Control Tool for Different Accelerometer Data*, München, Institut für Medizinische Statistik und Epidemiologie, 2013

Tragezeit für die auszuwertende Datei berechnet werden. Dies konnte mit dem Button *wearing Time* (siehe Abb. 1: 6) mit Hilfe des Algorithmus von Hecht et al 98 durchgeführt werden. Resultierend geht aus diesem Schritt ein mit 0 oder 1 befüllter Vektor hervor. Dabei steht 0 für nicht getragen und 1 für getragen. Aufzufinden war der Vektor sowohl als MATLAB-Datei *wearingTime.mat* als auch als csv-Datei *wearingTime.csv* vorzufinden. Die csv-Datei bestand aus einer Zeitspalte sowie einer Spalte mit den zugehörigen, errechneten Werten.

Für weiteres Interesse verweisen wir auf den Praktikumsbericht *Quality Control Tool for Different Accelerometer Data* von Johannes Eder und Martin Attenkofer.

5 Das erweiterte Tool

Bei der Erweiterung des gerade vorgestellten QualityControlTools kam es, neben den Änderungen in der Funktionalität, auch zu einigen Änderungen in der Oberfläche. Diese werden im Folgenden genauer erläutert. Im Anschluss daran folgt eine detaillierte Erklärung der dahinterstehenden Funktionalitäten und Implementierungen.

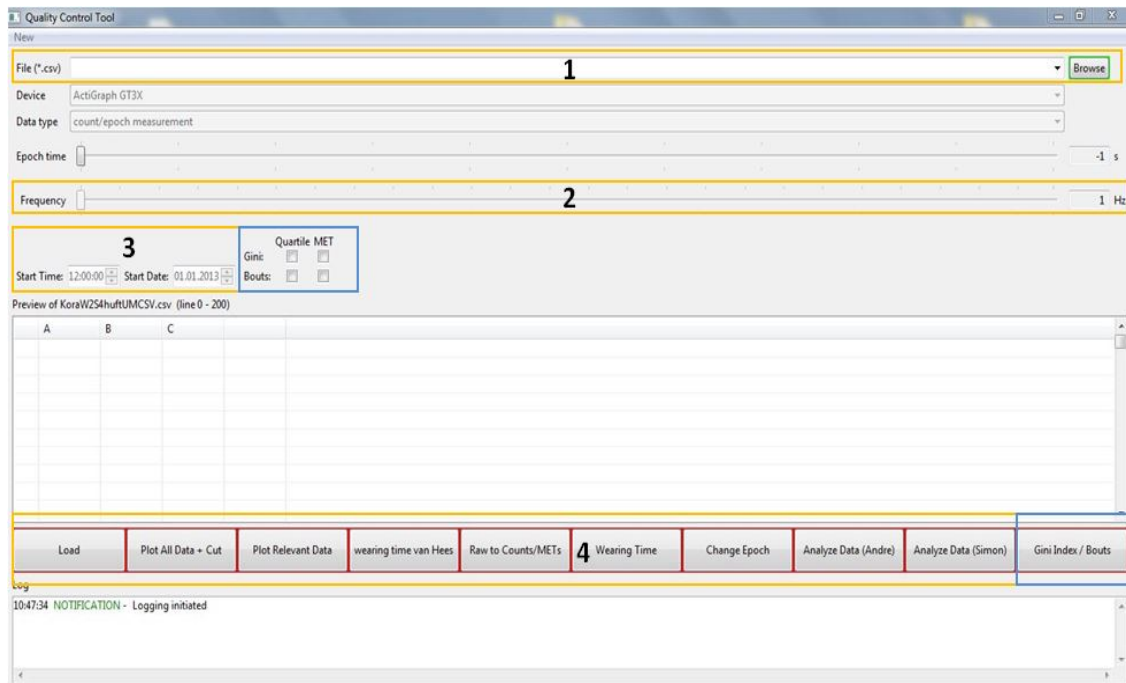


Abbildung 2: GUI des QualityControlTools 2. Version

In Abbildung 2 sind alle Änderungen zur vorherigen Version gekennzeichnet. Dabei wird in dieser Arbeit auf die blau markierten Felder nicht eingegangen, da diese Gegenstand einer anderen, parallel laufenden, Arbeit sind.

In Feld 1 der Abbildung 2 hat sich für den Betrachter nichts geändert. Allerdings verbirgt sich hinter dem *Browse*-Button eine andere Funktionalität. Statt die Datei zu laden wie in der ersten Version des Tools, wird hier nur noch die Datei ausgewählt und der entsprechende Dateipfad in die JAVA-Umgebung geladen. Das eigentliche Laden in die MATLAB-Umgebung findet erst mit drücken des *Load*-Buttons (siehe Abb. 2: 4) statt. Da die wesentlichen Änderungen des Tools darin liegen, dass mehrere Sensorarten und auch Rohdaten ausgewertet werden können musste die Funktionalität des *Load*-Buttons ebenfalls erweitert werden. Die genaue Dokumentation dafür ist in Kapitel 5.2 beschrieben.

Wie gerade erwähnt kann das erweiterte Tool auch Rohdaten auswerten. Diese haben keine Epochenlänge in Minuten oder Sekunden wie *count/epoch measurement*, sondern werden mit einer bestimmten Frequenz aufgenommen. Da nicht in allen Rohdaten diese Frequenz angegeben ist, muss der User diese wissen und per Hand eingeben. Dies ist vor Allem bei Shimmer und Somnowatch Daten nötig. Bei ActiGraph GT3X+ und GeneActive Daten wird die Frequenz automatisch ausgelesen (vgl. Kapitel 5.2). Jedoch sollte der User hier immer die Frequenz überprüfen. Damit der Anwender also Zugriff auf die Frequenz hat, diese einstellen und überprüfen kann ist die Leiste *Frequency* (siehe Abb. 2: 2) hinzugekommen.

In der ersten Version des Tools mussten die auszuwertenden Daten immer um 12:00 Uhr beginnen. Da die jetzigen Daten auch zu anderen Zeitpunkten anfangen können wurde das Tool um die Felder *StartTime* und *StartDate* erweitert (siehe Abb. 2: 3). Diese muss der User für Somnowatch und Shimmer Daten per Hand eingeben, da in den entsprechenden Dateien weder Informationen über Anfangszeit noch über Anfangsdatum gespeichert sind. Für Daten der anderen Sensoren werden diese Informationen automatisch eingelesen (vgl. Kapitel 5.2).

Im Bereich der Buttons (siehe Abb. 2: 4) sind ebenfalls einige Neuerungen gemacht worden. Zum einen wurde der oben erwähnte *Load*-Button neu erstellt. Zum Anderen wurde für die Auswertung von Rohdaten der Button *wearing Time van Hees* hinzugefügt (vgl. Kapitel 5.5). Um es dem User zu ermöglichen Rohdaten in Epochdaten umzuwandeln wurde zudem der Button *Raw to Counts/METs* neu angelegt (vgl. Kapitel 5.6). In den nachstehenden Kapiteln wird nun auf die Implementierung der einzelnen Buttons näher eingegangen.

5.1 Der Browse-Button

5.1.1 Aktion des Benutzers

Um mit dem QualityControlTool arbeiten zu können, muss zunächst der *Browse*-Button betätigt werden. Im Zuge dessen muss die zu analysierende csv-Datei aus dem entsprechenden Ordner auf dem Computer des Users ausgewählt werden.

5.1.2 Aktionen des QualityControlTool

Ist die auszuwertende Datei gewählt, stellt das QualityControlTool vorläufige Berechnungen an. Es wird aus der csv-Datei der Name des jeweiligen Sensors ausgelesen und in die GUI des Tools eingetragen. Anhand der Datei wird ebenfalls die Art der Daten eingetragen. Handelt es sich um Rohdaten, erscheint auf der Oberfläche im Feld *Data type* der Eintrag *raw data measurement*, andernfalls *count/epoch measurement*. Zusätzlich wird die Frequenz, sofern es sich um Rohdaten handelt, der Sensordaten bestimmt und ebenfalls dem Benutzer

mittels der Frequenzleiste dargelegt. In Abbildung 3 sind diese geänderten Einstellungen in der Oberfläche anhand einer Beispieldatei dargestellt.

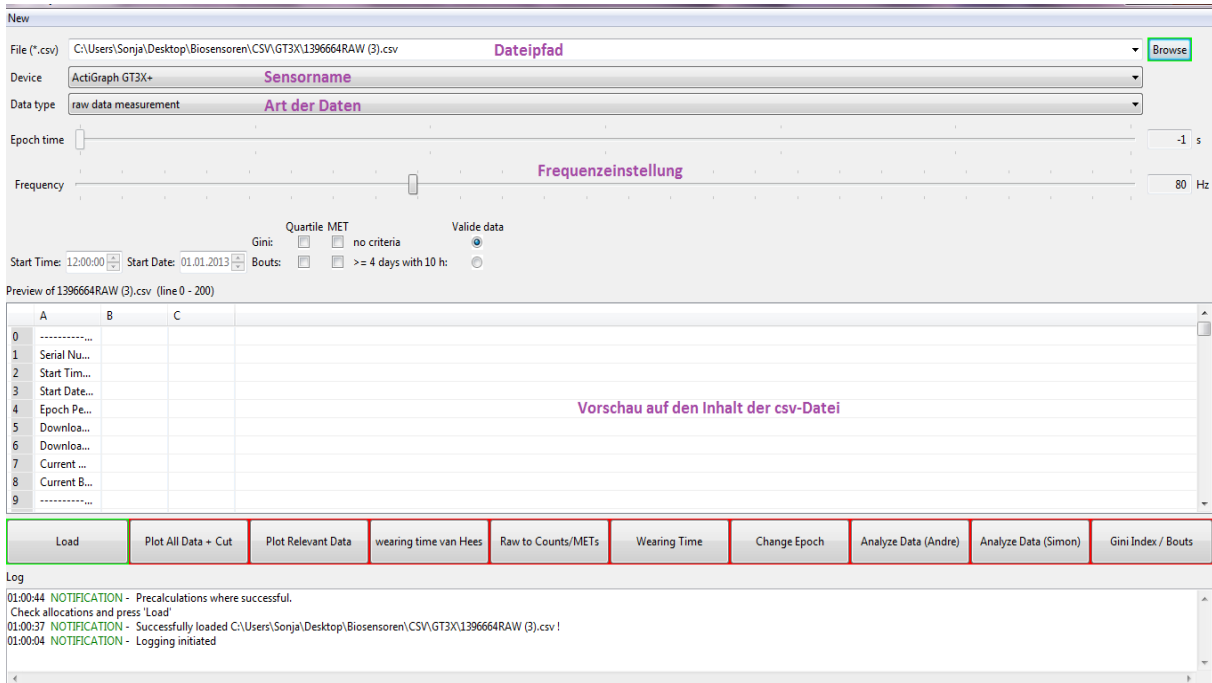


Abbildung 3: GUI nach Betätigen des *Browse*-Buttons

5.1.3 Implementierung

Im JAVA-Code werden in der Klasse *KoraSteps* die Voreinstellungen nach Aktivierung des *Browse*-Button bestimmt. Dies findet in der Methode *public static boolean precalculation(String currentPath, String currentFile, String currentDevice)* statt. Hier wird auf die entsprechenden MATLAB-Funktionen zugegriffen.

Den MATLAB-Funktionen wird dabei der Dateipfad und der Dateiname der zuvor ausgewählten csv-Datei übergeben. Es wird die Funktion *FindSensorName(1, file_path)* aufgerufen. Entsprechend des Funktionsnamens liest diese aus der Kopfzeile der Datei den Namen des behandelten Sensors. Die Variable für den Namen des Senors wird schließlich in JAVA auf den ermittelten Namen gesetzt und die GUI dementsprechend eingestellt. Misslingt die Suche, behandelt JAVA dies mit einem Auffangen der geworfenen Exception. Zusätzlich wird der User mit der Mitteilung “Unknown Device Type” benachrichtigt, was den Benutzer darauf hinweist den Sensortyp per Hand richtig einzustellen.

Die Frequenz des Sensors wird mit Hilfe der MATLAB-Funktion *FindSensorFrequency*(1, *file_path*) bestimmt. Auch dieses Ergebnis aktualisiert in JAVA die aktuelle Variable für die Frequenz und ist für den Benutzer auf der GUI ersichtlich gemacht. Ein möglicherweise eintretender Fehler wird wiederum mit einer *Exceptionbehandlung* abgefangen und dem User mit der Nachricht "Unknown Frequency!" kenntlich gemacht. Auch hier sollte der Benutzer diese Einstellung per Hand vornehmen.

Erfolgreiches Abschließen aller Vorberechnungen wird mit der Meldung "Pre-calculations where successful. Check allocations and press "Load"" bestätigt.

Außerdem erscheint eine Vorschau der csv-Datei auf dem Display der Benutzeroberfläche. Dafür ist, wie auch schon in der ersten Version des Tools, die Funktion *private void setupCsvTable(CsvFileLoader csvFile)* in der JAVA-Klasse *MainControlsBackend* zuständig.

Zum Schluss wird der *Load*-Button zur Nutzung freigeschaltet. Dies wird dem Nutzer durch eine grüne Umrandung des Buttons signalisiert.

5.2 Der Load-Button

Der *Load*-Button wird dem Benutzer nur nach dem Ausführen des *Browse*-Vorgangs freigeschaltet.

5.2.1 Möglichkeiten für den Benutzer

Die Trennung des *Browse*-Buttons und des *Load*-Buttons erlaubt dem User vom Programm selbstständig vorgenommene Einstellungen zu überprüfen und gegebenenfalls zu verifizieren. So lassen sich nach den Voreinstellungen sowohl der Sensorname, als auch die Frequenz, die Datenart und Startzeit und -datum einstellen.

5.2.2 Implementierung

Die ausgewählte Datei wird nun mit den endgültigen Einstellungen erstmals in die MATLAB-Umgebung geladen. In der JAVA-Klasse *KoraSteps* wird die Methode *KoraStep1* ausgeführt. Abhängig von dem jeweiligen Sensor und dem *measurement* werden von hier aus die entsprechenden MATLAB-Funktionen aufgerufen. Für das *count/epoch measurement* existiert in diesem Tool nur eine Implementation für den Sensor ActiGraph GT3X. Für die Sensoren Shimmer, GeneActive und ActiGraph GT3X+ gibt es hingegen lediglich eine Implementierung für *raw data measurement*. Im Falle von falschen User-Einstellungen im Bezug auf die zugrundeliegende Datenart wird dies als Fehler behandelt. Dabei erscheint die Meldung "No code implemented for [Sensorname] count/epoch measurement!" beziehungsweise "No code implemented for ActiGraph GT3X raw data measurement".

Für jeden dieser Fälle existiert eine eigene MATLAB-Funktion. Diese weichen in Sensor- und Datentypspezifischen Details voneinander ab. Grund dafür

ist der Unterschied im Aufbau der csv-Dateien. So besitzen beispielsweise die GeneActive Dateien einen Dateikopf mit den jeweiligen Informationen wie Messfrequenz oder Startzeit, wohingegen Somnowatch Dateien direkt mit den gemessenen Beschleunigungsdaten starten. Auch die Anzahl der Spalten weichen voneinander ab.

In Matlab wird in der Funktion *FileLoader*_*[Sensorname]Kora* bzw. für *count/epoch measurement* *FileLoaderKora.new* die Messdaten aus der csv-Datei gefiltert und zeitlich angepasst. Das Ergebnis wird unter einer Datei mit dem Namen *[Filename]_RAW.mat* bzw. *[Filename].mat* gespeichert und ist so für den User jederzeit wieder aufrufbar.

Ist das Laden abgeschlossen, wird sowohl der *PlotAllData + Cut*-Button als auch der *RawToCounts/METs*-Button grün umrandet und somit zur Benutzung freigegeben.

5.3 PlotAllData + Cut -Button

Die geladenen Daten lassen sich in einer Grafik darstellen. In der Grafik sind für Rohdaten drei verschiedene Graphen zu sehen (siehe Abb. 4). Die dazugehörige Legende beschreibt, welche Bewegungsrichtung der jeweilige Graph darstellt. Zusätzlich ist die Grafik mit dem Sensorname, dem Datentyp, der Startzeit und der Frequenz beschriftet.

An der Darstellung des Plots für Counts hat sich nichts wesentliches im Vergleich zur ersten Version des Tools verändert.

In dem Fenster des *Plots* kann der Benutzer nun Einstellungen bezüglich Startzeit und Dauer der relevanten Daten treffen.

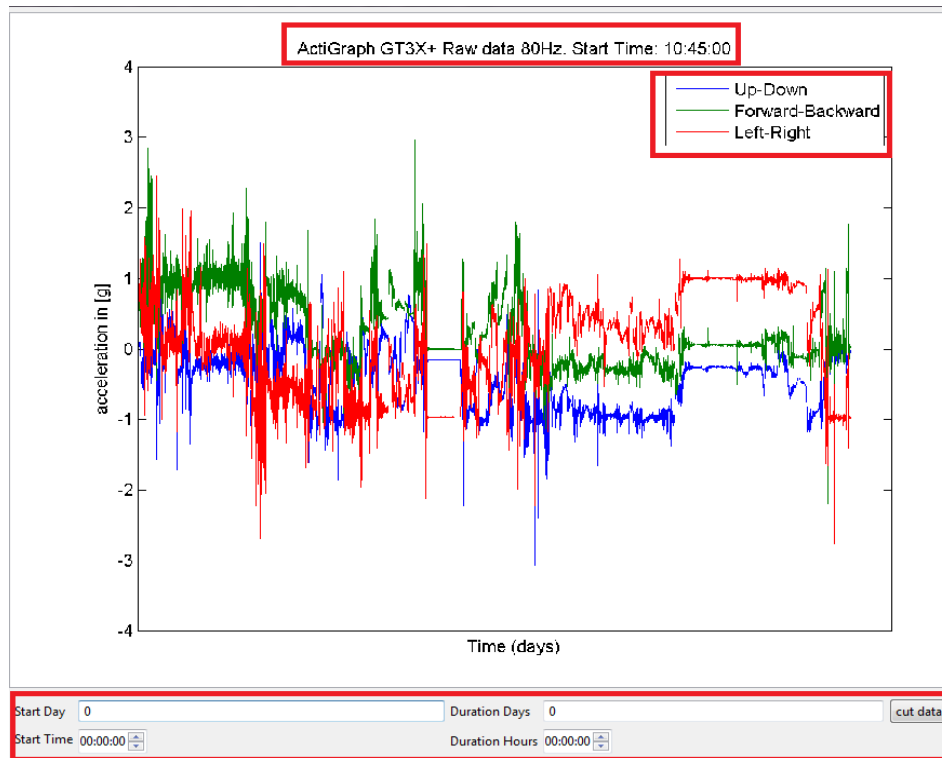


Abbildung 4: Beispiel: *Plot* einer GT3X+ Datei

5.3.1 Implementierung

Auch bei dem *Plotten* wird zunächst von der *KoraSteps.class* auf MATLAB zugegriffen. Von *step2_plotFile* wird je nach *measurement* *cutFirst1_Raw* bzw. *cutFirst1_new* aufgerufen.

Im Fall von *count/epoch measurement* berechnet die Funktion lediglich ein Schaubild zur einfacheren Visualisierung der Daten, wie schon in der ersten Version des Tools. Die einzigen Änderungen zur vorherigen Version sind eine ausführlichere Titulierung des erstellten Diagramms und eine Anpassung der Beschriftung auf der x-Achse. Auf dieser sind sowohl bei *count/epoch measurement* als auch bei *raw data measurement* die aufgezeichnete Zeit dargestellt. Dabei wird ein Tick immer genau um 12:00 Uhr eines Tages eingezeichnet. Da die Daten im Unterschied zur ersten Version aber auch zu anderen Zeiten starten können musste hier eine zusätzliche Berechnung eingefügt werden.

Bei *raw data measurement* zeigt das Diagramm statt nur einer Achse alle drei gemessenen Achsen. Diese werden, neben allen anderen Berechnungen, ebenfalls in der Methode *cutFirst1_Raw*, abhängig von dem jeweiligen Sensortyp,

beschriftet (vgl. Abb. 4).

5.3.2 Cut Data

Wie oben erwähnt gibt es bei den auszuwertenden Daten verschiedene Startzeiten. Damit der User möglichst flexibel ist, was das abschneiden der irrelevanten Daten angeht wurden die zwei Felder *Start Time* und *Duration Hours* hinzugefügt. Neben der schon vorhandenen Funktion im nächsten Schritt nur bestimmte Tage zu plotten, können jetzt auch bestimmte Zeiten ausgewählt werden. So ist es beispielsweise möglich nur die Daten eines Tages von 6:00 Uhr bis 17:00 Uhr zu betrachten.

Ändert der Benutzer nichts an den Einstellungen der Felder *Start Time* und *Duration Hours* so werden die Daten jeweils vom gewünschten Tag ab 00:00 Uhr bis zum letzten, gewünschten Tag um 00:00 Uhr dargestellt. Der letzte Tag berechnet sich dabei aus dem Eintrag im Feld *Start Days* plus dem Eintrag im Feld *Duration Days*. Überschreitet dieser Wert den vom Sensor gemessenen Zeitraum werden alle Daten ab dem gewünschten Startdatum dargestellt.

Falls das Feld *Duration Days* ebenfalls nicht geändert wird und somit der Eintrag auf 0 bleibt bedeutet dies, dass die auszuwertenden Daten eine Länge von 0 Tagen besitzen. Das führt in weiteren Berechnungen zu Fehlern, da kein Datensatz mehr zur Verfügung steht. Deshalb sollte der User immer darauf achten, beim beschneiden der Daten eine sinnvolle und gültige Zeitangabe zu machen.

5.4 PlotRelevantData-Button

Der *PlotRelevantData-Button* wird dem Benutzer erst nach dem Ausführen des *Plots* aller Daten und dem Schneidevorgang zur Verfügung gestellt. Als Ausgabe erscheint die Grafik der zugeschnittenen, relevanten Daten. Das Fenster entspricht dem im Kapitel 5.3 vorgestellten. Allerdings lässt sich diese Grafik hier nicht weiter schneiden. Demnach lässt dieses Fenster dem Benutzer keine weiteren Angaben zur Dauer bzw. Startzeit zu.

5.4.1 Implementierung

Wird der Button von dem Benutzer betätigt, wird im JAVA-Programm die Methode *KoraStep3* der Klasse *KoraSteps* aufgerufen. Von hier aus wird die MATLAB-Funktion *step3_cutFile* geladen. In dieser gibt es dann die Fallunterscheidung für *raw measurement* und *count/epoch measurement*. Im ersten Fall wird jeweils die MATLAB-Funktion *cutFirstPlus2_Raw*, im zweiten Fall die Funktion *cutFirstPlus2_new* berechnet. Die in der Datei *[Filename]_RAW.mat* bzw. *[Filename].mat* gespeicherten Daten werden dabei in einem neuen Graphen gezeichnet. Der so entstehende, zweite Plot wird wiederum in *[Filename]_SecondPlot.png* beziehungsweise in

[*Filename*]*_SecondPlot_RAW.png* abgespeichert. Somit lässt sich auch diese Grafik von dem Benutzer ohne erneute Einstellungen jederzeit aufrufen.

5.5 Wearing time van Hees-Button

Da die Berechnung der Tragezeit nach Hecht nur für Counts möglich ist wurde zusätzlich der Button *Wearing time van Hees* angelegt. Dieser stellt dem Benutzer die Möglichkeit zur Berechnung der Tragezeiten bei *raw data measurement* zur Verfügung. Hierbei wird der Algorithmus von van Hees verwendet. Der Button wird freigeschaltet sobald der Schritt *PlotAllData + Cut* erfolgreich beendet wurde.

5.5.1 Implementierung

Sobald der User den Button gedrückt hat wird über den JAVA-Code die Methode *public static boolean KoraStep5VanHees(String device, String file, int frequency)* in der Klasse *KoraSteps* aufgerufen. Über diese wird dann wiederum die MATLAB-Funktion *wearingTimevH_Raw(sensor, file, hz)* geöffnet und berechnet. Zur Berechnung werden die Daten aus der Datei [*Filename*]*_RAW* verwendet. Das heißt, wurden die irrelevanten Daten bereits mit der Funktion *PlotRelevantData* abgeschnitten, fehlen diese auch in der *wearingTime*-Berechnung. Wurde dieser Schritt (vgl. Kapitel 5.4) bisher noch nicht durchgeführt wird die Tragezeit für alle eingelesenen Daten ausgewertet. Dabei betrachtet der Algorithmus die eingehenden Daten immer in Intervallen von 15 Minuten. Aus diesem Grund müssen die auszuwertenden Daten auch eine Länge von mindestens 15 Minuten aufweisen, andernfalls wird der Fehler "Error while performing van Hees Wearing Time calculation! Data too small!" ausgegeben.

Das Ergebnis wird auf zwei verschiedene Arten abgespeichert (siehe Abb. 5): Es wird eine *.mat*-Datei erstellt. In dieser befindet sich eine Matrix *B* die in den ersten drei Spalten die gemessenen Beschleunigungsdaten des Sensors enthält und in der vierten Spalte jeweils die Information, ob der Sensor laut van Hees getragen wurde oder nicht. Wurde der Sensor getragen ist der Eintrag gleich 1, wurde er nicht getragen ist der Eintrag gleich 0.

csv-Datei				mat-Datei					
	A	B	C		1	2	3	4	5
1	05-Oct-2011 14:30:00,1			1	0.1328	0.0786	0.1158	1	
2	05-Oct-2011 14:45:00,1			2	0.1577	0.0693	0.1389	1	
3	05-Oct-2011 15:00:00,1			3	0.2324	0.1528	0.1881	1	
4	05-Oct-2011 15:15:00,1			4	0.1242	0.0554	0.1641	1	
5	05-Oct-2011 15:30:00,1			5	0.0908	0.0570	0.1418	1	
6	05-Oct-2011 15:45:00,1			6	0.2763	0.1120	0.2277	1	
7	05-Oct-2011 16:00:00,1			7	0.3407	0.0953	0.2326	1	
8	05-Oct-2011 16:15:00,1			8	0.1690	0.0477	0.1244	1	
9	05-Oct-2011 16:30:00,1			9	0.3234	0.0807	0.1924	1	
10	05-Oct-2011 16:45:00,1			10	0.3702	0.0807	0.2675	1	
11	05-Oct-2011 17:00:00,1			11	0.4308	0.1058	0.2887	1	
12	05-Oct-2011 17:15:00,1			12	0.1794	0.1441	0.1704	1	

Abbildung 5: Auswertung der Tragezeit nach van Hees

Außerdem wird eine csv-Datei abgespeichert. Diese enthält die genaue Zeitangabe jedes berechneten Intervalls und die dazugehörige Trage-Information aus der vierten Spalte der Matrix *B*. Zu finden sind diese Dateien jeweils unter dem Namen *[Filename]_wearingTimevHees* mit der entsprechenden Endung für den Datentyp.

5.6 Raw to Counts/METs-Button

Sobald eine Datei mit dem *Load*-Button (vgl. Kapitel 5.2) in die MATLAB-Umgebung geladen wurde und es sich bei den Daten um *raw data measurement* handelt besteht die Möglichkeit diese in *count/epoch measurement* umzuwandeln. Diese Funktionalität stellt der Button *Raw to Counts/METs* für den Benutzer zur Verfügung.

5.6.1 Aktionen des QualityControlTool

Der Button wird aktiv, das heißt seine Umrandung wird grün, sobald der *Load*-Vorgang erfolgreich abgeschlossen ist und solange im Feld *Data type* der Typ *raw data measurement* ausgewählt ist. Nachdem die Berechnungen durch diesen Button erfolgreich abgeschlossen wurden wird das Feld *Data type* automatisch auf *count/epoch measurement* gesetzt und alle damit verbundenen Aktionen werden freigegeben. Damit ist der *Raw to Counts/METs*-Button nicht mehr aktiv, also für den User nicht mehr anwählbar.

5.6.2 Implementierung

Drückt der Benutzer den *Raw to Counts/METs*-Button wird in der JAVA-Klasse *KoraSteps* die Methode *public static boolean RawDataToEpoch(String currentPath, String currentFile, String currentDevice, int currentFrequency, String startDate, String startTime)* aufgerufen. Diese öffnet sogleich die MATLAB-Funktion *calculateCounts1(path, file, hz, sensor, d, t)*. Da für die Berechnung

der Counts die Daten nochmals direkt aus der csv-Datei gelesen werden wird hier, wie auch bei dem *Load*-Button (Kapitel 5.2), eine Fallunterscheidung zwischen allen Sensortypen gemacht.

In jedem dieser Fälle werden zuerst die Daten aus der angegebenen Datei gelesen und in eine Matrix gespeichert. Dies geschieht jeweils mit einem Sensor-spezifischen Funktionsaufruf. Mit dieser Matrix werden in einem zweiten Schritt jeweils die Counts/METs berechnet. Diese Berechnung ist für die Sensoren GeneActive, Shimmer und Somnowatch gleich und wird in der MATLAB-Funktion *rawDataToEpoch3*(*A*, *hz*, *DateTimeNum*, *file*) durchgeführt. Für den Sensor ActiGraphGT3X+ muss die Berechnung angepasst werden. Deshalb wird in diesem Fall nur die Funktion *rawDataToEpoch2*(*path*, *file*, *hz*) aufgerufen. Da diese Funktion also nur für den Sensor ActiGraphGT3X+ verwendet werden kann werden in ihr auch gleich die Daten aus der csv-Datei eingelesen. Das heißt für diesen Sensor gibt es keinen extra Funktionsaufruf zum Einlesen der Daten in eine Matrix.

Bei jedem Sensor werden die Rohdaten automatisch in Counts mit einer Epochenlänge von 60 Sekunden umgewandelt. Die Ergebnisse werden dann auf zwei verschiedene Arten abgespeichert: Als erstes werden sie als TIME-SERIE unter dem Namen *[Filename].mat* gespeichert. Mit dieser Datei ist es dann möglich alle weiteren Berechnungen durchzuführen, die auch schon bei der ersten Version des Tools für *count/epoch measurement* implementiert waren.

Außerdem werden die Counts und METs in einer csv-Datei unter dem Namen *[Filename]_Count_MET.csv* gespeichert. Diese Datei enthält dann zuerst die genaue Zeitangabe der Daten. Dahinter stehen jeweils, durch Kommata getrennt, die zugehörigen Counts und dann METs.

Damit nach der erfolgreichen Umwandlung die Oberfläche auch das Richtige anzeigt wird im JAVA-Code als letzter Schritt in dieser Berechnung die Epochen-Skala aktualisiert und auf 60 Sekunden gesetzt.

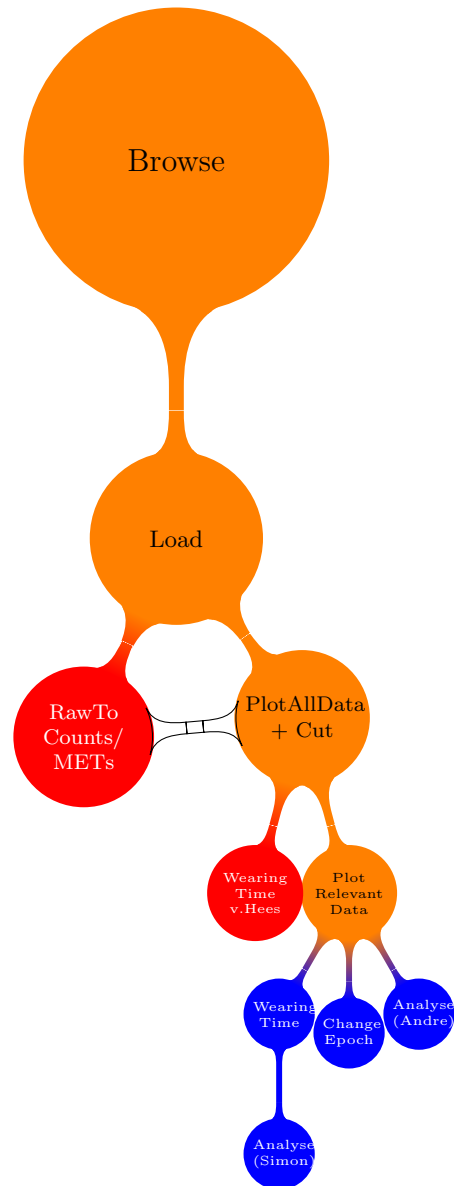
5.7 Analyse nach Andre

5.7.1 Beschreibung

Bei Betätigung des *Analyse (Andre)*-Button werden die Daten nach dem Algorithmus von Andre ausgewertet. Pro Tag der Eingabedaten werden eine bestimmte Anzahl an Werten berechnet. Die einzige Änderung in dieser Berechnung zur ersten Version des Tools ist, dass die Werte nun unter dem Namen *[Filename]_finaldata.csv* gespeichert werden (siehe Abb.6).

“QualityControlTool for Different Accelerometer Data” von Martin Attenkofer und Johannes Eder.

6 Abhängigkeiten im QualityControlTool



Legende:
Abhängigkeiten von großen zu kleinen *Bubbles*

rot $\hat{=}$ *raw data measurment*
blau $\hat{=}$ *count/epoch measurment*
orange $\hat{=}$ *raw data measurment* oder *count/epoch measurment*

7 Beispielablauf einer Analyse von Shimmer Daten

Im Folgenden werden die Funktionen des erweiterten Tools anhand der Auswertung einer Datei veranschaulicht. Dazu dient eine Datei des Sensors Shimmer. Es wird nicht mehr auf die genaue Implementation eingegangen, sondern lediglich alles beschrieben, was der User für die Benutzung des Tools benötigt. Auf die Schritte der Veränderung der Epoche und die Analysen nach Simon sowie nach Andre wird nur kurz eingegangen, da sich hier zur Vorgängerversion nichts wesentliches verändert hat.

7.1 Starten des Programms

Der Aufruf des Programms funktioniert ganz einfach über einen Doppelklick auf die ausführbare jar-Datei. Sollte das nicht funktionieren, kann das Programm über einen Doppelklick auf die Batch-Datei alternativ geöffnet werden.

Damit das Tool auf dem Rechner funktioniert müssen dort einige Vorinstallationen gemacht werden. Neben der Software die bereits für die erste Version des Tools² benötigt wird, wird jetzt auch die **MATLAB Signal Processing Toolbox** benötigt.

²Johannes Eder, Martin Attenkofer, *IDP: Quality Control Tool for Different Accelerometer Data*, München, Institut für Medizinische Statistik und Epidemiologie, 2013

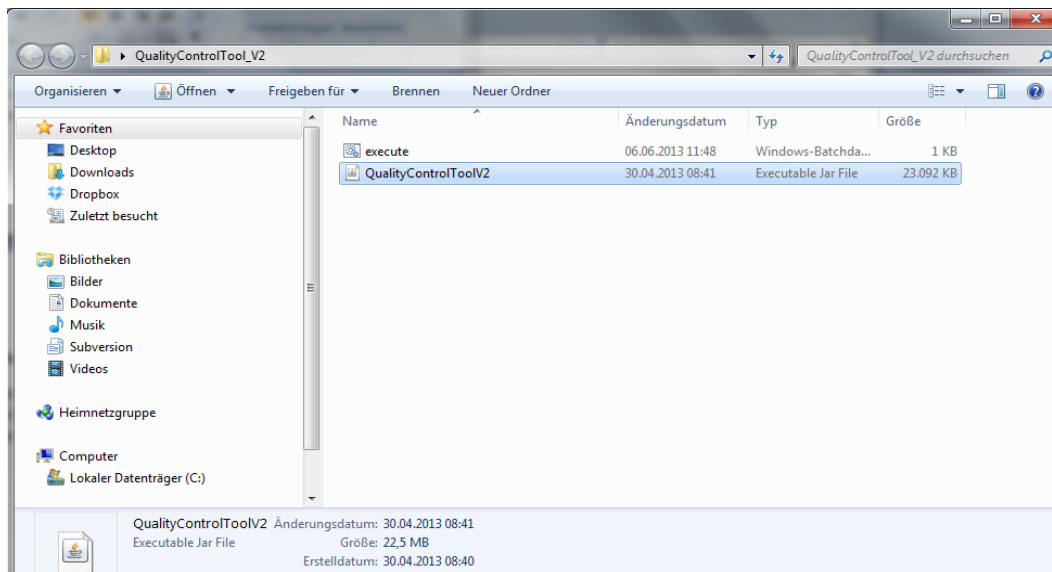


Abbildung 7: Öffnen des Tools

7.2 Laden einer Shimmer Datei

Zuerst wird über den *Browse*-Button (siehe Kapitel 5.1) auf dem Computer eine auszuwertende Datei ausgewählt. In diesem Fall handelt es sich um eine Datei mit Rohdaten von dem Sensortyp Shimmer.

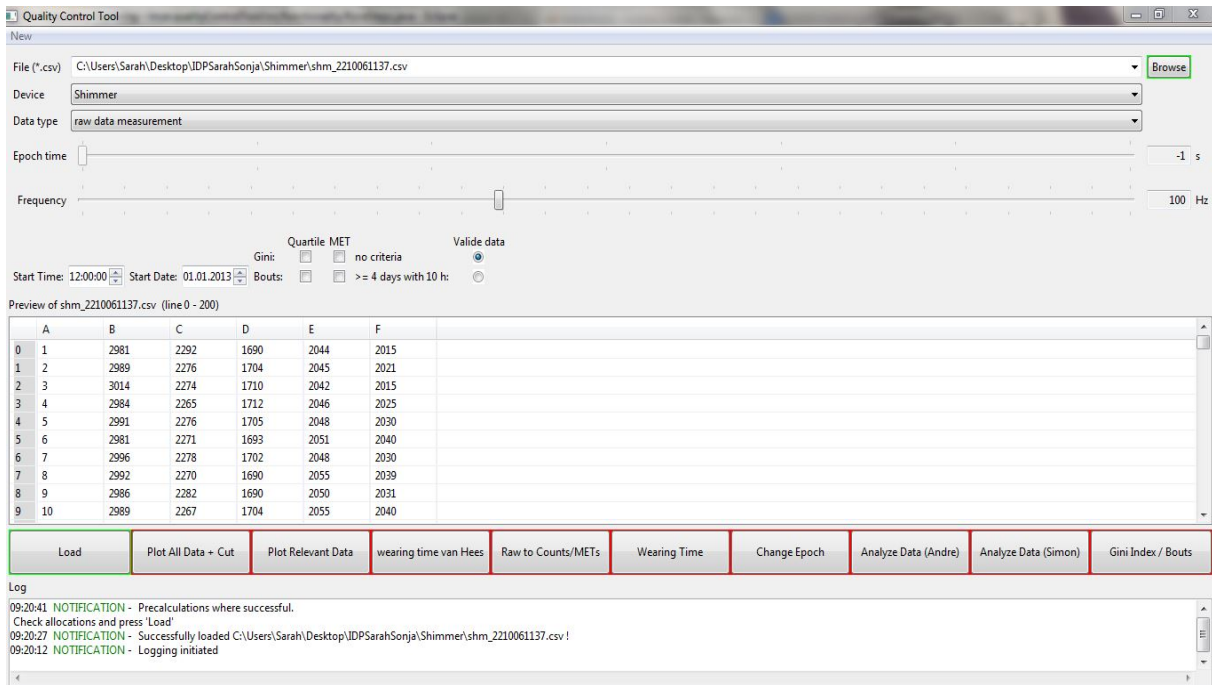


Abbildung 8: Nach auswählen der Datei mittels *Browse*-Button

Diese Informationen werden direkt aus der Datei ausgelesen und erscheinen nach den Vorberechnungen auf der Oberfläche in den jeweiligen Feldern (siehe Abbildung 8). Somit muss der Nutzer keine Informationen über die Datei haben. Allerdings sollte trotzdem immer überprüft werden ob die Einstellungen wirklich stimmen, da es ansonsten zu Problemen oder falschen Berechnungen führen kann.

Ebenfalls wird die Frequenz automatisch gesetzt. Da diese jedoch nicht in den Daten des Sensortyps Shimmer zu finden sind wurde hier ein Defaultwert von 100 Herz angegeben. Diesen Wert gilt es für den Benutzer unbedingt zu überprüfen, da er auch abweichen kann.

Eine weitere Besonderheit die bei den beiden Sensortypen Shimmer und Somnowatch zusätzliche zu beachten ist, sind die Felder *Start Date* und *Start Time*. Auch diese beiden Informationen sind in den Daten der genannten Sensoren nicht enthalten. Deshalb müssen diese vor drücken des *Load*-Buttons vom Benutzer unbedingt angepasst werden. Geschieht dies nicht, so wie in diesem Beispielaufbau, wird jeweils der voreingestellte Defaultwert verwendet.

Hat der Benutzer alle Einstellungen überprüft und gegebenenfalls geändert kann nun der grün umrandete *Load*-Button betätigt werden (siehe Kapitel 5.2). Wur-

de dieser gedrückt kann es vor Allem bei größeren Daten, abhängig von der jeweiligen Rechnerleistung, zu einer Rechenzeit von mehreren Minuten kommen.

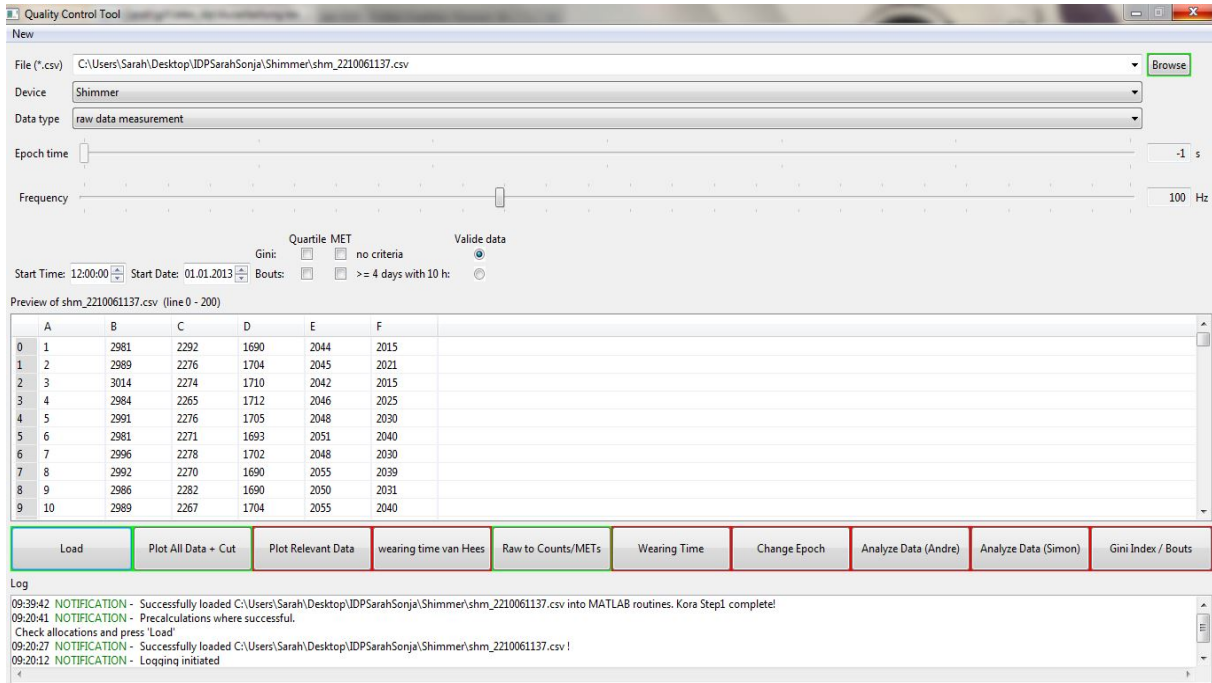


Abbildung 9: Nach laden der Datei mittels *Load*-Button

7.3 Umwandeln von Rohdaten in Counts

In diesem Schritt des Beispielablauf werden nun die Counts aus den Rohdaten berechnet. Natürlich besteht die Möglichkeit zuerst die Rohdaten zu Plotten (PlotAllData + Cut-Button) und zu beschneiden. Allerdings ist es für große Daten sinnvoll zuerst die Counts zu berechnen. Mit diesen dauert die Berechnung des ersten Plots nicht so lange und die Daten können auf die selbe Art beschnitten werden. In den folgenden Abschnitten wird beschrieben wie der Benutzer danach über den *Data type* zurück zu *raw data measurement* wechseln kann um nur noch den Plot für die ausgewählten Rohdaten durchzuführen.

Um die Rohdaten in Counts umzuwandeln muss der User lediglich den, nach dem Laden grün umrandeten, Button *Raw to Counts/METs* (siehe Kapitel 5.6) drücken. Nach der Berechnung wird in der Oberfläche die Auswahl des *Data type* automatisch auf *count/epoch measurement* gesetzt. Nun ist es möglich zwischen beiden Einstellungen hin- und herzuspringen um Berechnungen durchzuführen die es eventuell nur für einen der Datentypen gibt.

Ebenfalls wird nun die Leiste für die Einstellung der Epoche freigegeben und auf 60 Sekunden gesetzt. Gleichzeitig kann der Benutzer nun keine Änderungen mehr in der Frequenz-Leiste mehr machen (vgl. Abb. 10).

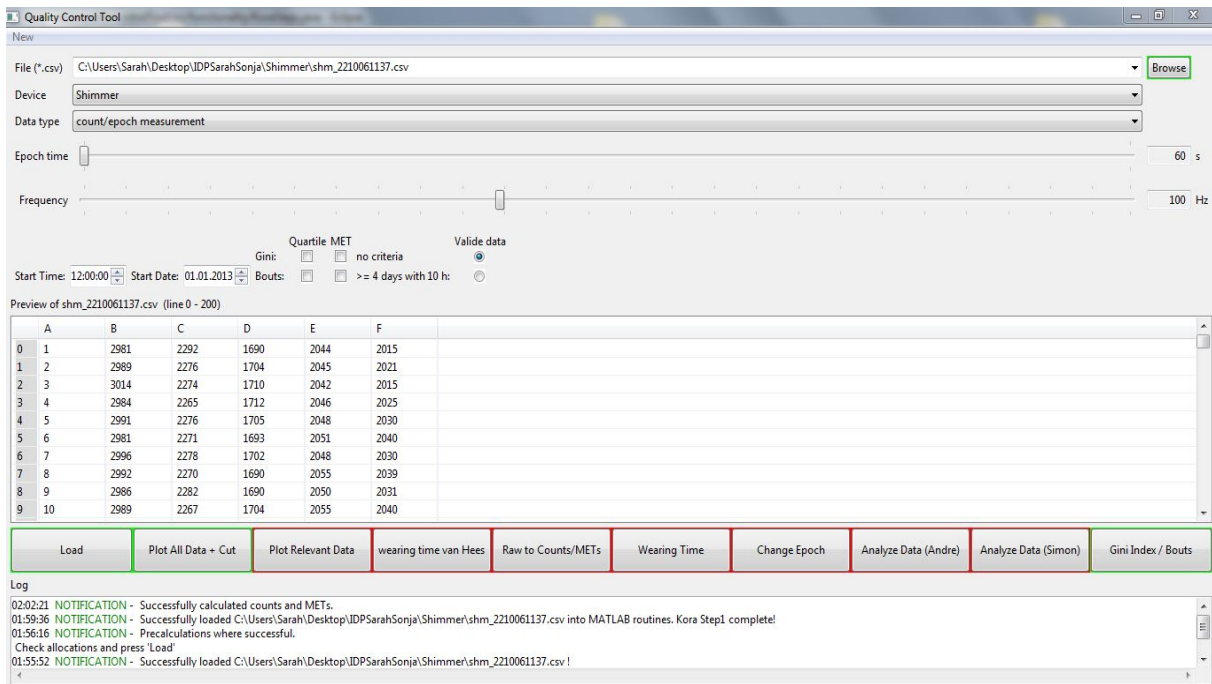


Abbildung 10: Nach berechnen der Counts/METs mittels *Raw to Counts/METs*-Button

7.4 Plotten aller Daten

Um nun die Daten auf einen wesentlichen Teil zu konzentrieren und unwichtige Teile abzuschneiden werden in diesem Schritt erst alle Daten in einer Grafik dargestellt. Anhand dieser Grafik kann der Benutzer dann ein Zeitintervall angeben in dem die wichtigen Daten liegen. Ausschließlich die Daten in diesem Zeitintervall werden dann für weitere Berechnungen benutzt.

Zur Erstellung der Grafik für den ersten Plot wird der *PlotAllData + Cut*-Button gedrückt (siehe Kapitel 5.3). Für die Beispieldaten sieht dieser wie folgt aus:

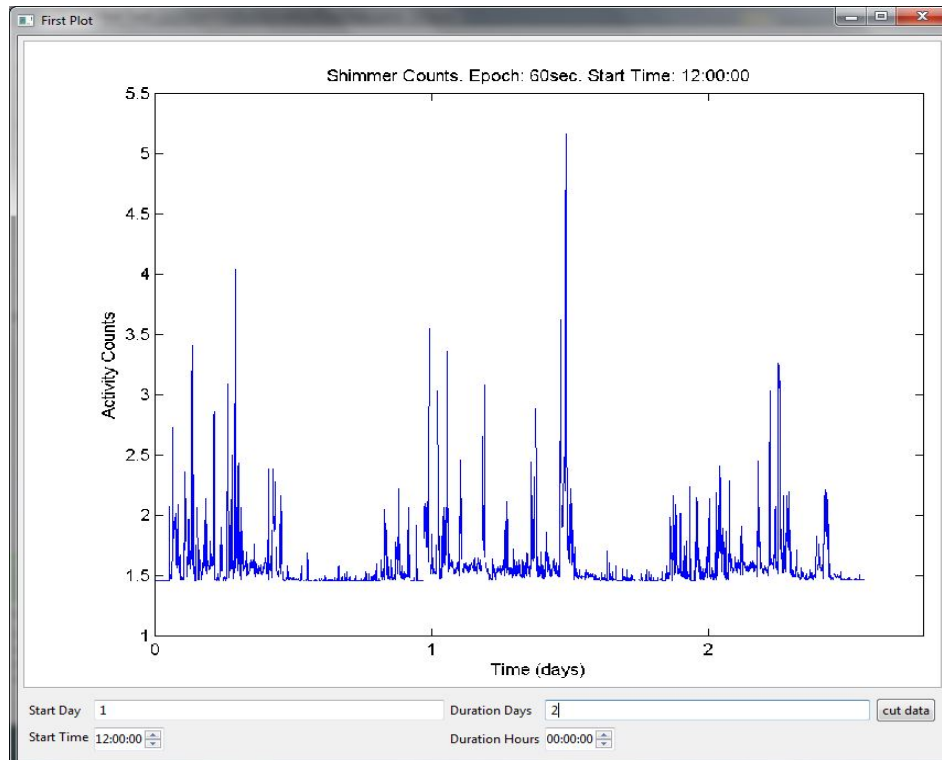


Abbildung 11: Erster Plot

Im unteren Bereich wird nun das Zeitintervall für die zukünftig verwendeten Daten angegeben. In diesem Beispiel wird der Bereich zwischen dem ersten und dem dritten Tag genutzt. Dies ergibt sich aus der 1 im Feld *Start Day* und der 2 im Feld *Duration Days*. Da die Daten in diesem Fall schon vor Ende des dritten Tages aufhören, werden in den folgenden Schritten automatisch alle Daten vom ersten Tag an bis zum Ende verwendet. Um die Auswahl zu bestätigen muss vom Anwender der *cut data*-Button in dem Plot Fenster gedrückt werden.

Wird vom User nun der Button *PlotRelevantData* angewendet werden nur die gerade gewählten Daten in einer weiteren Grafik dargestellt (siehe Abb. 12). Erst ab diesem Schritt werden auch für die Analyse und die Berechnung der Tragezeit nur noch die gewählten Daten verwendet.

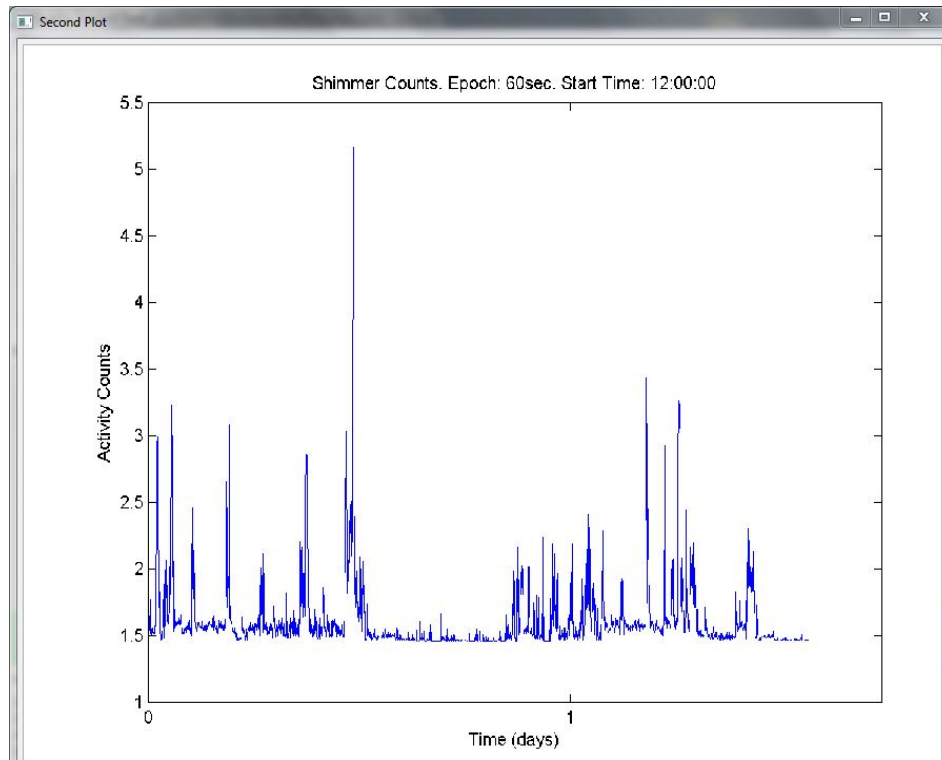


Abbildung 12: Zweiter Plot

7.5 Plotten der relevanten Rohdaten

Damit die im vorherigen Schritt ausgewählten Daten für die Rohdaten genauso beschnitten werden muss der Button *PlotRelevantData* gedrückt werden nachdem in dem Feld *Data type* der Typ *raw data measurement* vom Nutzer eingestellt wurde (vgl. Abb. 13).

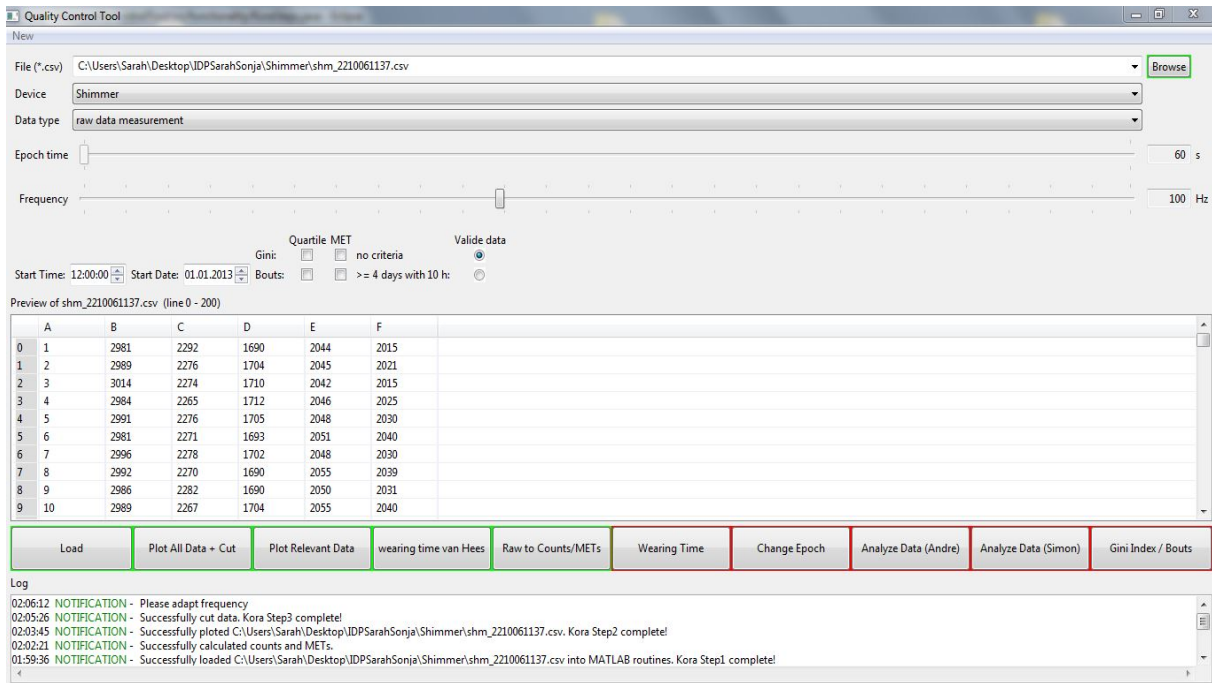


Abbildung 13: Ansicht der Benutzeroberfläche vor Plotten der relevanten Rohdaten

Da jetzt ein Teil der Daten wegfällt funktioniert dieser Schritt schneller als der erste Plot mit Rohdaten, der wie in Kapitel 7.4 beschrieben, umgangen wurde. Dennoch kann es auch hier zu Rechenzeiten von mehreren Minuten kommen. Sind alle Berechnungen abgeschlossen erscheint die Grafik für den zweiten Plot mit Rohdaten:

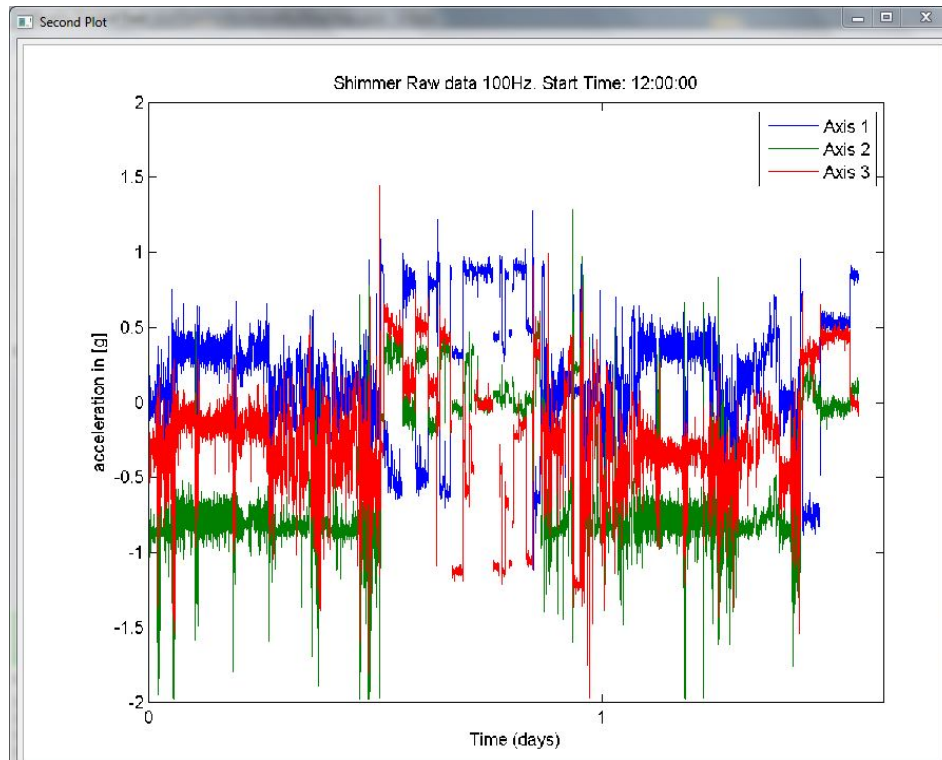


Abbildung 14: Zweiter Plot mit Rohdaten

Nun werden für folgende Anwendungen auch bei Rohdaten nur noch die gewählten, nicht mehr alle, Daten verwendet.

7.6 Berechnung der Tragezeit nach van Hees

In dem nächsten Schritt kann nun die Tragezeit nach van Hees berechnet werden. Da sich diese ausschließlich mit Rohdaten berechnen lässt werden für die Auswertung die gerade beschnittenen Daten verwendet.

Damit dies durchgeführt wird muss der User lediglich auf den *Wearing time van Hees*-Button drücken. Beachtet werden sollte davor noch, wie gerade erklärt, dass dieser nur funktioniert solange der *Data type* noch auf *raw data measurement* gesetzt ist.

Nach der Berechnung sieht die Oberfläche wie folgt aus, und der Benutzer bekommt über den *Log* eine Bestätigung der erfolgreichen Berechnung mitgeteilt:

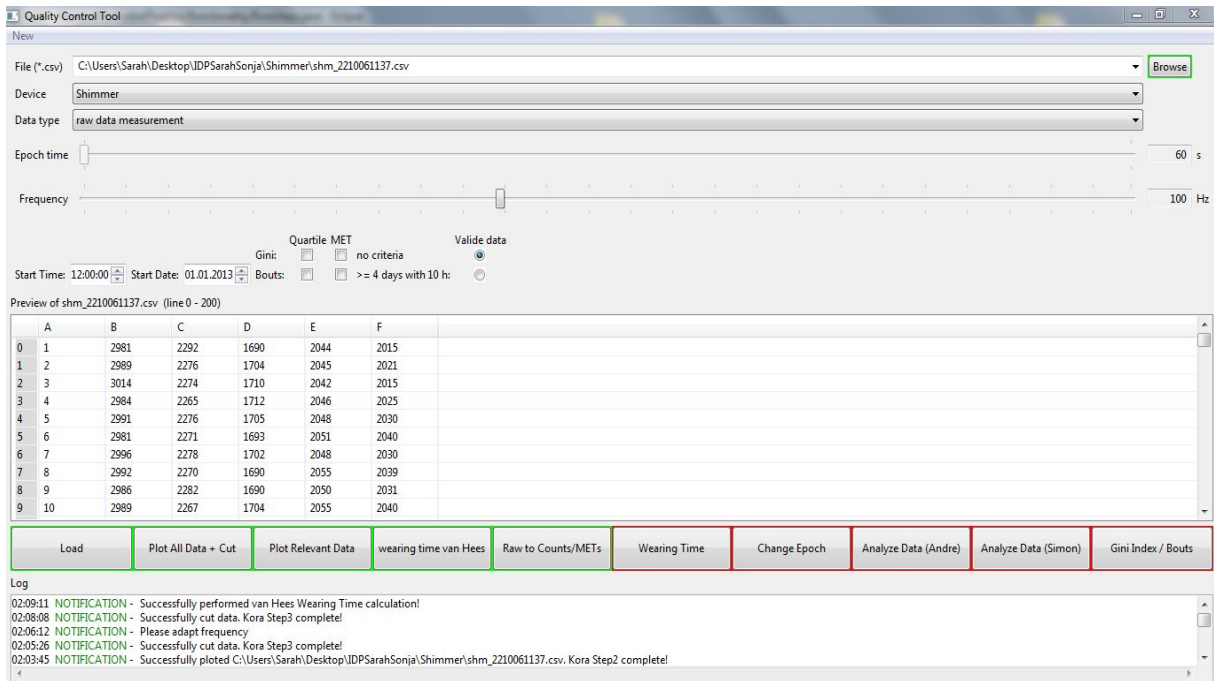


Abbildung 15: Nach der Berechnung der Tragezeit nach van Hees

Die Ergebnisse sind, wie in Kapitel 5.5 erklärt, in diesem Beispiel unter den Namen *shm_2210061137_wearingTimevH.csv* und *shm_2210061137_wearingTimevH.mat* zu finden. Im Gegensatz zu den zwei Plots, werden diese Ergebnisse nicht automatisch geöffnet. Um sie anzeigen zu lassen muss der User in den Ordner des QualityControlTools navigieren. Dort sind die gerade erwähnten Dateien dann zu finden und können ganz normal gehandhabt werden.

7.7 Epochenänderung

An dem Schritt zur Änderung der Epoche wurde in der Erweiterung keine Neuerungen vorgenommen. Mit Hilfe des Schiebereglers (siehe Abb. 16) kann eine beliebige Epoche eingestellt werden, sofern der *Data type* wieder auf *count/epoch measurement* gesetzt wurde.

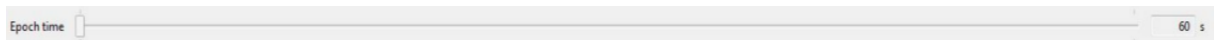


Abbildung 16: Schieberegler zur Epocheneinstellung

Das Intervall zur Anpassung beginnt bei der von dem Sensor festgelegten initialen Epochenlänge (beispielsweise 2 Sekunden). Das Maximum liegt bei 60

Sekunden. Die Wahl der neuen Länge muss ein vielfaches der initialen Epochenlänge betragen.

7.8 Analyse nach Andre

Die Daten aus der Analyse nach Andre werden in der csv-Datei `[Filename]_finaldata.csv` gespeichert. Wie auch in der vorherigen Version kann es bei dem Öffnen der Datei in *LibreOffice* bzw. *OpenOffice* zu Fehlern bei der Fließkommadarstellung kommen. So wird eine Fließkommazahl teilweise als ganze Zahl dargestellt. Für weitere Erklärungen verweisen wir auf den Praktikumsbericht der 1. Version.

Bei der Auswertung werden allein Tage mit einer Tragedauer des Sensors von mindestens 10 Stunden berücksichtigt.

7.9 Analyse nach Simon

Die Analyse nach Simon lässt sich erst nach der Berechnung der Tragezeit durchführen. Somit muss noch ehe der Button für die Analyse betätigt werden kann, der *WearingTime*-Button gedrückt worden sein. Die Zeiten sind in einer csv-Datei hinterlegt und können bei der Analyse ausgelesen werden.

Da bei der Erweiterung dieser Schritt nicht bearbeitet wurde, lassen sich nur die *count/epoch measurement* Daten des Sensors Actigraph auswerten.

8 Hinweise zur Benutzung des Tools

Das QualityControlTool läuft auf jedem Computer. Allerdings müssen zunächst die Umgebungen **Java Runtime Environment**³, **Matlab Compiler Runtime**⁴ und die **MATLAB Signal Processing Toolbox**⁵ installiert werden.

Allerdings gibt es bei der Programmausführung je Computer Unterschiede in der Berechnungsgeschwindigkeit. Dies liegt an der Größe des Arbeitsspeichers eines Computers und an der Anzahl der vorhandenen Kernels. Entsprechend der Kernelanzahl werden die Berechnungen in den MATLAB-Funktionen auf die Kernels verteilt. Je mehr Kernels existieren und belastet werden, desto schneller ist auch das Programm. Dennoch können Berechnungen – vor allem von großen Datenmengen – viel Zeit in Anspruch nehmen. Aus diesem Grund empfehlen wir während der Ausführung des Tools anderweitige, gleichzeitige Benutzung des Rechners zu vermeiden bzw. auf das Nötigste zu beschränken. Dazu zählt beispielsweise auch das Ausschalten des Internets. Dadurch kann der Rechner für das Programm vollkommen ausgelastet werden und dieses

³mögliche Downloadseite: <http://www.oracle.com/technetwork/java/javase/downloads/index.html>

⁴mögliche Downloadseite: <http://www.mathworks.de/products/compiler/>

⁵mögliche Downloadseite: <http://www.mathworks.de/products/signal/>

gewinnt an Geschwindigkeit.

Zudem gibt es für jeden Sensor Besonderheiten, die beachtet werden müssen:

	Shimmer	Somnowatch	GeneActive	ActiGraph GT3X+
Load	Zeit, Datum und Frequenz einstellen		Einstellungen überprüfen	
Raw to Counts	Falls danach Wechsel zu Rohdaten, Frequenz neu einstellen			

9 Mögliche Verbesserungen und Erweiterungen

Um das QualityControlTool zu verbessern und zu erweitern weisen wir auf folgende Anregungen hin:

- Da es besonders bei großen Daten und Computern mit vergleichsweise wenig Arbeitsspeicher zu Problemen mit dem Einlesen kommen kann, würde das Ersetzen der MATLAB-Funktion *csvread* eventuell zu einem schnelleren Ablauf führen. Denn eigentlich werden aus den Dateien nur die drei Spalten für die Beschleunigungsdaten benötigt, allerdings werden im Moment alle gefüllten Spalten einer Datei eingelesen. Das heißt es wird viel Rechenaufwand darauf verwendet Daten in die MATLAB-Umgebung zu laden die im Anschluss für keine Funktion mehr benötigt werden.
- Zudem könnte die GUI mit einem zusätzlichen Balken versehen werden, der dem Benutzer anzeigt, wie viel Zeit eine Berechnung noch benötigt. Dies könnte mit einer Prozentangabe realisiert werden, die angibt wie viel Wartezeit abhängig von der Gesamtdauer noch einzurechnen ist.
- Eine weitere Einstellungsmöglichkeit die implementiert werden kann, ist eine Variable Epochenlänge bei der Umwandlung von Roh- zu Countdaten. Im Moment besteht nämlich nur die Möglichkeit Counts mit einer Epochenlänge von 60 Sekunden aus Rohdaten zu erstellen.
- Um dem Benutzer das Auswerten von vielen Daten des selben Sensors zu erleichtern könnte man in weiteren Arbeiten eine Batch-File erstellen. Diese könnte beispielsweise alle Dateien aus einem gewählten Ordner automatisch nacheinander auswerten, ohne dass der Anwender das Tool nach der Ordnerwahl weiterhin bedienen muss.
- Außerdem gibt es noch weitere, zahlreiche Funktionen für Analysemöglichkeiten der KORA Studie, die in das Tool mit eingebunden werden können. Ein Beispiel hierfür ist die Berechnung und geeignete Darstellung des Gini Index.

10 Fazit

Das QualityControlTool erleichtert den Umgang mit den MATLAB -Funktionen. So benötigt der User keine Kenntnisse über das Programmieren mit MATLAB und kann sich allein auf die Auswertungen konzentrieren.

Um dies zu realisieren, wurden zunächst die MATLAB-Funktionen angepasst. Viele Funktionen wurden für alle Sensoren zusammengefasst. Die bestehenden aus der 1.Version wurden an das *raw data measurement* sowie an die verschiedenen Sensoren angepasst bzw. erweitert (vgl. Kapitel 5). Dabei wurde die grundlegende Hierarchie der Vorgängerversion beibehalten und lediglich um Zusatzschritte erweitert. Mit Hilfe des MATLAB eigenen Deploytools wurden diese Funktionen als JAVA-Klassen verfügbar gemacht.

Anschließend wurden in JAVA die entsprechenden Erweiterungen implementiert, in denen die jeweiligen MATLAB-Funktionen aufgerufen werden. Dazu wurde die Oberfläche entsprechend angepasst. Durch das Verbinden der Buttons in der Benutzeroberfläche mit den Funktionalitäten, entstand ein einfach benutzbares Programm. Dies lässt sich an dem in dieser Ausarbeitung dargestellten Beispielablauf (siehe Kapitel 7) erkennen.

Es ist ein Programm entstanden, das das Auswerten auf einer großen Bandbreite ermöglicht. Dazu benötigt der Benutzer keinerlei Hintergrundinformationen im Bezug auf die Programmierung. Allerdings benötigt es sehr viel Geduld bei der Analyse. Besonders bei großen Dateien bedarf es für die einzelnen Schritte mehr oder weniger viel Zeit.