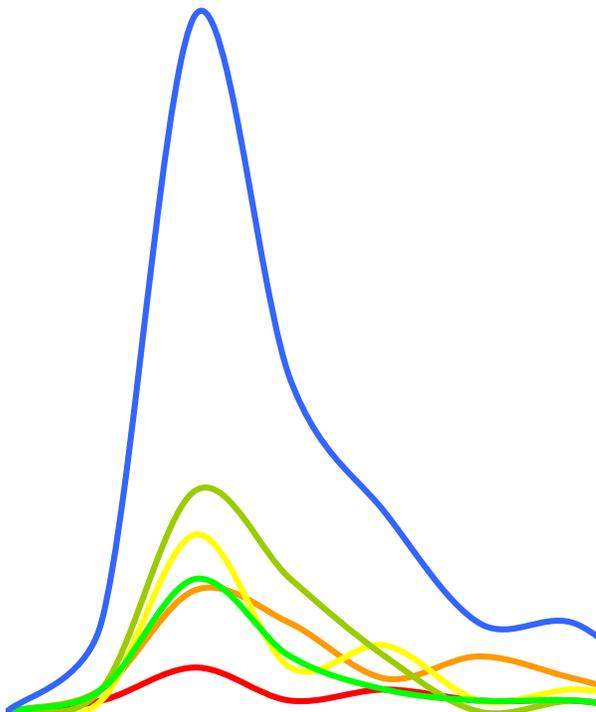




Universität Zürich
Institut für Informatik

Leistungssteigerung durch Context-Awareness

-
Aufbau, Implementierung
& Durchführung
eines Experiments



Dynamic and Distributed
Information Systems

Diplomarbeit 30. November 2007

Roger Trösch

von Etziken SO, Schweiz

Matrikelnummer: 01-718-774

rogertr@troesch.net

Betreuer: **Peter Vorburger**

Prof. Abraham Bernstein, PhD
Institut für Informatik
Universität Zürich
<http://www.ifi.uzh.ch/ddis>

Danksagung

An dieser Stelle möchte ich mich bei allen Personen herzlich bedanken, die mich bei der Erstellung dieser Arbeit unterstützt haben. Ein ganz besonderer Dank gilt meiner Familie, die mir das Studium ermöglicht und mich dabei unterstützt hat.

Weiter möchte ich mich bei Professor Abraham Bernstein bedanken, der mir diese spannende Arbeit ermöglicht hat. Bei Peter Vorbürger, meinem Betreuer, möchte ich mich auch ganz herzlich für seine engagierte Unterstützung bedanken.

Ein spezieller Dank gilt auch Felix-Robinson Aschoff, Stefan Christiani, Marcel Steffen & Stefan Weibel, sowie allen anderen Diplomanden, für ihre Hilfestellungen in allen Fragen und Sorgen. Natürlich dürfen auch die Testpersonen, die beim Experiment mitgemacht haben, nicht vergessen werden. Ohne sie, wäre diese Arbeit nicht zustande gekommen.

Zusammenfassung

Diese Arbeit ist im Gebiet Context-Awareness angesiedelt. Es soll dabei herausgefunden werden, ob mit einem mobilen Gerät die Arbeitsleistung einer Person erhöht werden kann. Um dies zu überprüfen, wurde ein Experiment entwickelt, implementiert und durchgeführt. Das Ziel ist es, einen möglichen Zusammenhang zwischen dem Kontext einer Situation und der Leistungsfähigkeit zu finden, welcher dann zur Steigerung der Arbeitsleistung eingesetzt werden könnte.

Als technische Grundlage dient ein mobiles Gerät, welches die Situation und deren Umgebung (Kontext) mit verschiedenen Sensoren erfasst. Mit einem psychologischen Test wurde die Konzentrations- und Leistungsfähigkeit der Probanden in Abhängigkeit zu ihrem Kontext ermittelt. Das Experiment wurde über einen Zeitraum von vier Wochen mit fünf Probanden durchgeführt. Anschliessend wurden die Daten ausführlich analysiert und ausgewertet. Zum Schluss, wurde der Effekt einer möglichen Leistungssteigerung der Probanden, anhand verschiedener Modelle abgeschätzt.

Abstract

This thesis is based on research within the domain of context-awareness. The aim is to determine whether it is possible to increase work performance by using a mobile device. An experiment was developed, implemented and conducted in order to verify a potential impact. The aim was to find a correlation between a person's environmental context and the work performance.

In the experiment sensors attached to a mobile device were used. These sensors recorded the context of a situation in which test subjects were doing a psychological performance test. By analysing the test results and the recorded context information, we tried to find coherence to predict the performance level in certain situations. Furthermore, we analysed these predictions and tried to estimate the effect of a potential work performance improvement.

Inhaltsverzeichnis

Inhaltsverzeichnis	ix
1 Einleitung	1
2 Vorgängige Arbeiten	3
2.1 Kommunikation und Arbeitsunterbrechung	3
2.2 Forschungsbereiche	5
2.2.1 Visuell und multimodalen Strategien	6
2.2.2 Temporal Strategies for Interruption	7
2.2.3 Event Percetion	7
2.3 Weitere Erkenntnisse	8
3 Das Experiment	11
3.1 Aufbau des Experiments	11
3.2 Psychologische Leistungstests	13
3.2.1 Konzentrations-Leistungstest	13
3.2.2 Aufmerksamkeits- und Belastungstest d2	16
3.2.3 Frankfurter Aufmerksamkeits-Inventar FAIR	17
3.2.4 Evaluation der Leistungstests	19
3.3 Technische Infrastruktur	21
3.3.1 Die Hardware	21
3.3.2 Die Software	25
3.4 Die Probanden	31
3.5 Durchführung des Experiments	32
3.5.1 KLT-Eingewöhnung und Abgabe der Geräte	32
3.5.2 Der Testbetrieb	33
3.5.3 Datenübertragung	34
3.5.4 Abschluss Trainingsphase	35
3.5.5 Produktivphase	35

4	Auswertung der Daten und Resultate	37
4.1	Die Aufbereitung der Daten	37
4.1.1	Die Aufbereitung der Sensordaten	37
4.1.2	Aufbereitung der KLT-Resultate	38
4.2	Deskriptive Statistik	40
4.2.1	Unterbrechungsstufe	43
4.2.2	Vorhersage der Leistungsfähigkeit mit Sensordaten	48
4.2.3	Interview mit den Probanden	56
5	Diskussion	57
5.1	Zusammenhang Unterbrechungslevel und Konzentrations- Leistungstest	57
5.2	Vorhersagen der Konzentrationsleistung anhand der Sensordaten	58
5.3	Weiterer Verlauf des Experiments	60
6	Einschränkungen und Ausblick	63
7	Schlussfolgerung	65
A	Anhang	67
A.1	Antworten aus Probandeninterview	67
B	Anhang	73
B.1	Benutzeranweisung für Experiment	73
C	CD	81
	Abbildungsverzeichnis	83
	Tabellenverzeichnis	85
	Literaturverzeichnis	87
	Literaturverzeichnis	87

1

Einleitung

Menschen haben das Bedürfnis jederzeit auf ihre Informationen zurückgreifen zu können, um jederzeit einen Informationsvorsprung zu haben. Sie wollen auch jederzeit mit anderen Leuten in Kontakt treten können. Vor einigen Jahren war dies noch ein Wunsch, den man nur schwer verwirklichen konnte. Heute ist dies mit wenig Aufwand umsetzbar und wir sehen es beinahe schon als Selbstverständlichkeit an. Die Flut an Medien- und Kommunikationsinhalten ist kaum mehr zu stoppen und wir benötigen bereits wirkungsvolle Massnahmen um daran nicht zu "ertrinken". Durch die Tatsache, dass man bereits jetzt und in Zukunft, noch viel stärker, zu fast jedem Zeitpunkt und an beinahe jedem Ort auf der Welt erreichbar ist, kann dies durchaus auch negative Auswirkungen haben. Mit der Einfachheit, mit der man eine Vielzahl von Mitteln rasch und effektiv für die Kommunikation einsetzen kann, wird beispielsweise eine Person öfters um Rat gefragt als noch vor einiger Zeit. Dieses vermehrte Aufkommen von Anfragen oder auch selbst getätigte Anrufe, können die Arbeitsleistung, die eine Person beim Arbeitgeber erbringen sollte, reduzieren oder sogar verunmöglichen.

Diese Arbeit zielt darauf ab, herauszufinden, ob ein System die Leistungsfähigkeit einer Person erhöhen kann. Dies soll dadurch erreicht werden, dass Störungseinflüsse, beispielsweise durch ein Telefonat, zu bestimmten Zeitpunkten nicht stattfinden, sondern herausgezögert werden. Sensoren sollen dabei messen, ob es im Augenblick gerade ein "guter" Zeitpunkt ist um eine Person zu stören, oder eine Unterbrechung negative Auswirkungen auf die Konzentration haben würde. Um herauszufinden ob ein solches System umsetzbar ist, haben wir in dieser Arbeit ein Experiment durchgeführt, das dies untersuchen soll.

Forschung, die sich mit diesem Thema befasst, zielt oftmals auf die Unterbrechbarkeit von Personen ab. Dabei ist nicht die Leistungsfähigkeit im Fokus der Betrachtung sondern nur das Gefühl, ob eine Unterbrechung störend oder durchaus willkommen ist. Diese Arbeit geht an dieser Stelle noch einen Schritt weiter und versucht die Auswirkungen einer Unterbrechung zu ergründen. Um im alltäglichen Umfeld unser Experiment, an einer Person umsetzen zu können,

mussten die benötigten Sensoren möglichst robust und leicht zu transportieren sein. Bei unserem Experiment wurden deshalb Sensoren eingesetzt, die an einem Mobiltelefon angebracht werden konnten, um möglichst realitätsnahe Messungen machen zu können und der Proband dabei wenig behindert wird.

Aufbau dieser Arbeit

Diese Arbeit ist in drei Hauptteile gegliedert. Im ersten Teil werden verwandte Arbeiten behandelt und verglichen. Im zweiten wird das durchgeführte Experiment mit den technischen und psychologischen Grundlagen erläutert. Im letzten Hauptteil werden die Resultate ausgewertet, analysiert und diskutiert.

2

Vorgängige Arbeiten

In diesem Kapitel werden unterschiedliche Arbeiten betrachtet die im Themenbereich der Unterbrechbarkeit von Menschen anzusiedeln sind. Dabei werden einige grundlegende Erkenntnisse aufgezeigt und unterschiedliche Forschungsbereiche vorgestellt.

2.1 Kommunikation und Arbeitsunterbrechung

Seit 1876 durch Alexander Graham Bell [19] das Telefon erfunden wurde, hat sich die Echtzeitkommunikation¹ grundlegend verändert. Es ist seit diesem Zeitpunkt möglich mit einer Person zu sprechen, die sich nicht mehr in physischer Nähe befinden musste. Bei einer normalen "Face-To-Face"-Kommunikation wird vor einem Gespräch, die Situation analysiert, in welcher sich die beiden Gesprächspartner befinden. Es kann dann von beiden Beteiligten entschieden werden ob ein Gespräch zu Stande kommen soll. Dabei werden Regeln angewendet, die uns bereits in der Kindheit beigebracht wurden. Diese werden meist implizit umgesetzt und lassen sich in folgenden Sätzen zusammenfassen [32]:

- **Wie wichtig ist es für mich, jetzt zu kommunizieren?**
- **Würde sich, aus meiner Sicht der Angesprochene daran stören unterbrochen zu werden?**
- **Wie ist die Beziehung zwischen den Kommunikationspartnern?**
- **Was für eine Art von Konversation soll zu Stande kommen? Für wen ist es wichtig? Wie lange wir sie dauern etc.?**
- **Ist es in dieser Situation sozial akzeptiert eine Konversation über ein bestimmtes Thema zu führen?**

¹Wir beschränken uns hier auf die elektronische Kommunikation und vernachlässigen die Übertragung von Schallwellen über weite Strecken z.B. mit einem Sprachrohrsystem auf einem Schiff.

Kommunizieren nun aber zwei Personen mit Hilfe von einem elektronischen Hilfsmittel, sind einige dieser fünf Regeln nicht ohne weiteres anwendbar. Es besteht die Möglichkeit, dass sich die angesprochene Person stark gestört fühlt und in ihrer Tätigkeit unterbrochen wird. Es herrscht somit ein Informationsdefizit der ansprechenden Person, welches durch die unterschiedlichen Lokalisationen der Beteiligten hervorgerufen wird. Vor einem Telefonanruf muss beispielsweise eine Annahme getroffen werden, in welcher Situation sich die anzurufende Person gerade befindet. Dies erklärt auch die weitverbreiteten Fragen am Anfang eines Telefonats, z.B.: "Wo bist du gerade? Was machst du gerade? Kannst du sprechen?" [32].

Obwohl die heutige Technologie bereits sehr fortschrittlich ist, gibt es noch fast kein Computer- oder Kommunikationssystem, das diese sozialen Konventionen mit entsprechendem Verhalten umsetzt. Meist ist es uns gar nicht bewusst, welchen großen Einfluss solche Systeme auf die soziale Situation haben. Das Klingeln eines Mobiltelefons während einer Besprechung oder die Meldung, dass ihr Akku jetzt vollständig geladen ist, beim Halten einer Präsentation, sind unangenehm, wenn nicht sogar peinliche Situationen, die durch die "Unwissenheit" von technischen Geräten verursacht werden. Dieser Effekt wird durch die Tatsache, dass solche Geräte immer kleiner und mobiler werden, noch zusätzlich verstärkt.

Um mit diesen unangenehmen Unterbrechungen umzugehen, gibt es nach J. Fogarty et al. zwei Strategien. Zum einen kann man versuchen keine proaktiven Systeme zu konstruieren oder sie nicht zu benutzen. Man müsste demnach ganz auf passiv wartende Systeme setzen, die nur durch die Interaktion des Benutzers funktionieren. Obwohl dieser Ansatz im Desktop-Computerbereich angemessen sein kann, profitieren mobile und Ubiquitous-Computersysteme² sehr stark davon [17].

Eine zweite Strategie wäre die temporäre Deaktivierung von potenziellen Störungsquellen, während ungünstigen Zeiten. Leider kann dieses Vorgehen in einem Eigentor enden, da beispielsweise ein ausgeschaltetes Mobiltelefon nicht nur die störenden, sondern auch die sehr wichtigen Anrufe zurückhält. Zusätzlich kann dann noch die Vergesslichkeit der Menschen hinzukommen, wodurch versäumt wird das Telefon, nach einer unpassenden Situation, wieder einzuschalten [33].

Aus einer simplen Unterbrechung können mehrere negative Folgeerscheinungen entstehen. Eine Studie der Nuklearindustrie [14] hat beispielsweise gezeigt, dass bei Vorfällen bei denen ein Atomkraftwerk heruntergefahren werden musste, 15% dadurch hervorgerufen wurden, weil das Überwachungspersonal von ihrer eigentlichen Arbeit abgelenkt worden ist [13].

Forschungen von R. Solingen im Bereich der Kommunikationsunterbrechungen haben gezeigt, dass 15-20% der Arbeitsleistung eines Arbeitnehmers dafür verwendet wird, um den Effekt einer Unterbrechung wieder wett zu machen. Dies sind etwa 15-20 Minuten pro Unterbrechung, die für die eigentliche Arbeit verloren gehen. Dieser Fakt hat die Wirtschaft schon längst erkannt und hat die Abkürzung COI (Cost of Interruption - Kosten der Unterbrechung) [23] eingeführt

²Ubiquitous Computing bezeichnet die Allgegenwärtigkeit von Computersystemen

[23]. Diese COI beinhalten alle relevanten Auswirkungen die eine Beeinträchtigung der Arbeitsleistung ausmachen und es gilt diese möglichst zu minimieren.

Eine Unterbrechung kann sehr viele unterschiedliche Ursachen haben, wobei nicht alle die gleich starken Auswirkungen nach sich ziehen. Neben den klassischen Hauptgründen für eine Unterbrechung, dem Telefon und dem "echten" Ansprechen, gibt es moderne proaktive Applikationen, die immer beliebter werden. Diese sind Email-Systeme [26], Instant Messaging [5] und Web Assistenten [25], welche die Aufmerksamkeit der Computerbenutzer sehr stark auf sich lenken können. Obwohl alle diese technischen Hilfsmittel einen positiven Nutzen bringen und aus den modernen Büros nicht wegzudenken sind, können sie ernstzunehmende Nebenwirkungen verursachen [1].

Neben der Tatsache, dass nach einer Unterbrechung die gestörte Person länger benötigt um die angefangene Arbeit wieder zu beginnen, konnten bei Personen noch weitere Effekte beobachtet werden. Unterbrechungen können beispielsweise den Stresslevel einer Person erhöhen, was zu einer emotionalen Belastung führen kann. Dies rührt daher, dass sie als lästig empfunden werden und Frustration ausgelöst wird, da man von der Arbeit abgehalten wird [24]. Dies führt dann zusätzlich zu einer Reduktion der Effizienz [1] [10].

Ein gutes System, dass die COI möglichst minimiert, sollte demnach an den genannten Punkten ansetzen und entsprechende Lösungen anbieten. Obwohl es im Grunde am einfachsten wäre, alle Kommunikationsmittel auszuschalten, ist dies in den meisten Fällen nicht anwendbar, und falls doch, nur für einen bestimmten Zeitraum möglich. Eine weitere Herausforderung, die an ein derartiges System gestellt wird, ist, dass es zwischen dringenden und weniger wichtigen Informationen diskriminieren sollte [1].

2.2 Forschungsbereiche

Wie bereits erwähnt, wird in dieser Arbeit versucht mit bestimmten Sensoren die Leistungsfähigkeit von Personen zu erhöhen. Viele Arbeiten, die im ähnlichen Umfeld angesiedelt sind benutzen für ihre Experimente keine mobilen Geräte, sondern arbeiteten mit festinstallierten Sensoren und Systemen. Es sollte zudem erwähnt werden, dass beinahe alle Experimente in diesem Forschungsbereich auf Laborexperimenten aufbauen und sich deshalb von realen Arbeitssituationen unterscheiden. Unser Experiment findet im realem Arbeitsumfeld statt und sollte so möglichst nahe an der Wirklichkeit sein. Zudem wird in den meisten Untersuchungen der Hauptaugenmerk auf die Unterbrechbarkeit gelenkt und nicht auf die Leistungsfähigkeit.

Die unterschiedlichen Forschungsbereiche im Gebiet der COI kann man in drei Gruppen un-

terteilen [1]. Es gibt die visuell und multimodalen Strategien³, die zeitlich⁴ und die Ereignis-sensiblen Strategien für die Unterbrechungsminimierung⁵, auf welche in den folgenden Abschnitten eingegangen wird.

2.2.1 Visuell und multimodalen Strategien

In diesem Bereich wird versucht mit Hilfe von Sensormodellen den optimalen Unterbrechungszeitpunkt zu finden. D. McFarlane [27] hat vier Taktiken für die Koordination von Unterbrechungen entwickelt. Diese sind: *immediate* (bei welcher der Benutzer sofort Antwort geben muss), *negotiated* (bei welcher der Benutzer wählen kann wann er antworten will), *mediated* (ein intelligenter Agent entscheidet über den besten Zeitpunkt) und *scheduled* (die Unterbrechungen kommen in festgelegten Intervallen, die im Voraus gewählt wurden). Seine Forschungsergebnisse haben aber gezeigt, dass keine dieser vier Methoden die absolut Beste ist. McFarlane folgerte daraus, dass der Benutzer selbst entscheiden sollte, welche Taktik für ihn am besten ist und er damit am wenigsten gestört wird.

Arroyo, Selker und Stouffs haben versucht herauszufinden welche Einflüsse aus der Umwelt am ehesten für eine Unterbrechung des Arbeitsflusses verantwortlich sein könnten. Sie fanden heraus, dass starker Geruch wie auch Vibration und Geräusche sehr stark ablenken können (siehe Abbildung 2.1) [34].

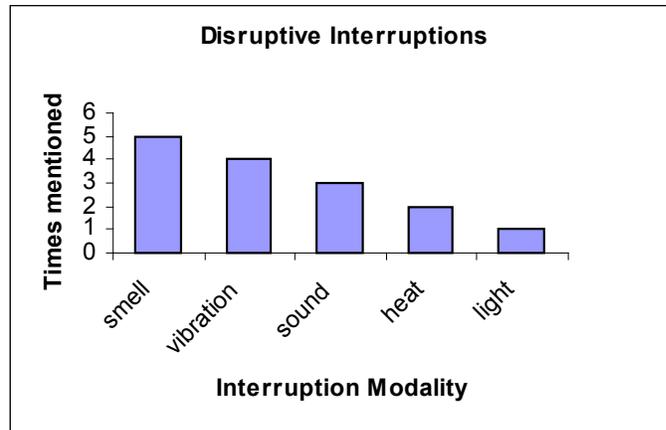


Abbildung 2.1: Unterbrechungsfaktoren [34]

Diese Arbeit hier würde auch in diese Kategorie (visuell und multimodalen Strategien) eingeteilt werden, da Sensoren für die Vorhersage eingesetzt wurden.

³Eng.: Visual and Multimodal Strategies for Interruption

⁴Eng.: Temporal Strategies for Interruption

⁵Eng.: Event Perception

2.2.2 Temporal Strategies for Interruption

Der zeitliche Aspekt für eine verlässliche Vorhersage spielt vor allem in der Arbeit von Miyata und Norman [28] eine wichtige Rolle. Sie haben herausgefunden, dass es drei Phasen bei der Bearbeitung von Aufgaben gibt. Diese sind *planning*, *execution* und *evaluation*. Eine Unterbrechung sollte nun jeweils am Anfang einer solchen Phase liegen, um die Zeit für die Wiederaufnahme möglichst kurz zu halten.

Eine weitere Arbeit, welche die zeitliche Einteilung einer Aufgabe in vier gleich grosse Quartile vorschlägt, hat die Resultate von Miyata und Norman bestätigt. Demnach wird es als angenehmer empfunden, wenn eine Unterbrechung zu Beginn einer Aufgabe auftritt. Dies könnte damit zusammenhängen, dass am Anfang einer neuen Tätigkeit noch genügend Zeit für die Bearbeitung zur Verfügung steht und diese dann kontinuierlich kürzer wird.

2.2.3 Event Percetion

In diesem Gebiet wird versucht zu ergründen, wie eine bestimmte Arbeit im Kopf hierarchisch zusammengesetzt ist. Dabei werden Einteilungen in grob und fein gemacht, wobei einfachere sich wiederholende Abläufe im Groben und präzise Abläufe im Feinen angesiedelt sind. Die Arbeit von P. Adamczyk und B. Bailey welche auf den Ideen von Temporal Strategies wie auch auf Event Percetion aufbaut, hat gezeigt, dass eine Unterbrechung von den Testpersonen besser aufgenommen wird, wenn sie auf einem hierarchisch hohen Ebene stattfindet (siehe Abbildung 2.2) [1].

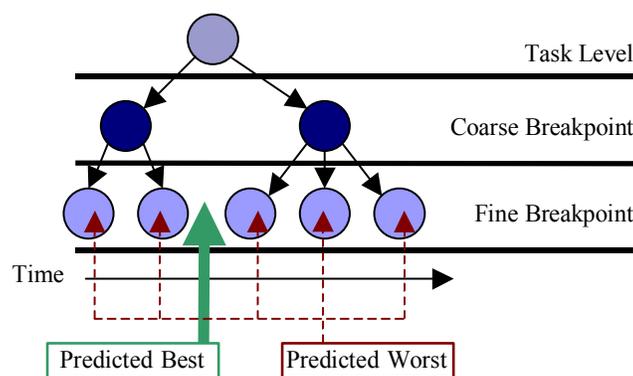


Abbildung 2.2: Hierarchische Aufteilung einer Arbeit [1]

2.3 Weitere Erkenntnisse

Weitere Arbeiten haben ergeben, dass eine Aufgabe, die durch eine gedanklich stark beanspruchende Störung unterbrochen wird, schwieriger wieder aufgenommen werden kann, als wenn eine kognitiv wenig fordernde Aufgabe zwischendurch bearbeitet wurde. Es konnte auch gezeigt werden, dass Instant Messages⁶, die für eine im Augenblick bearbeitete Aufgabe relevant sind, weniger störend wirken, als solche die irrelevant sind [6]. Hess und Detweiler haben zudem in einem Experiment herausgefunden, dass eine Aufgabe, die durch eine zweimalige Wiederholung geübt worden war und bei der dritten Wiederholung unterbrochen wurde, keinen positiven Effekt auf die Leistungsfähigkeit hatte, was vielleicht etwas erstaunt [8].

Man fand auch heraus, dass wenn die Anzahl der Unterbrechungen ansteigt, der Aufwand für die zu erledigende Arbeit steigt und der Wiedereinstieg in die ursprüngliche Arbeit unproportional grösser wird. Dieser Effekt lässt sich durch die gesunkene Leistungsfähigkeit, wie auch durch die gestiegene geistige Ermüdung erklären [24]. Um einer umfassenden Betrachtung gerecht zu werden, sollte zudem erwähnt werden, dass kurze Pausen von ein paar Minuten auch positive Effekte auf die Leistungsfähigkeit haben können. Es scheint aber, dass dies nur für kognitiv wenig anspruchsvolle Arbeit gilt, die einen stark repetitiven Charakter haben [15].

In der Arbeit von L. Dabbish und R. Kraut [7] wird nach einem Experiment, bei welchem die Koordination von Kommunikationspartnern untersucht wurde, festgestellt, dass die Kommunikation zwischen zwei Leuten am besten funktioniert, wenn der Initiator abstrakte Informationen über den Zustand der anzusprechenden Person bekommt. Es sollte beispielsweise in einem Computersystem permanent Informationen über die Arbeit der anderen Personen angezeigt werden, wobei alles Irrelevante ohne weiteres weglassen werden kann. Die Forscher hielten zudem fest, dass mehr Informationen über die aktuelle Tätigkeit nicht automatisch eine bessere Vorhersage bedeutet. Es muss auch darauf geachtet werden, dass die Privatsphäre der Personen nicht verletzt wird. Würden durch die Tatsache, dass der Tätigkeitsstatus für andere ersichtlich ist, unangenehme Gefühle entstehen, könnte dies sogar die Produktivität einschränken und den COI erhöhen, anstatt ihn zu senken.

In der Arbeit von E. Horvitz und J. Apacible [18] wird für eine Optimierung der Unterbrechungswirkung empfohlen, ein System zu verwenden, das mehrere verschiedene Informationsquellen kombiniert einsetzt. Diese können sich beispielsweise aus Sensordaten, Events auf dem Computer, Kalenderinformationen und Daten aus dem Arbeitsumfeld (visuelle und akustische) zusammensetzen. Mit Machine Learning könnte dann ein statistisches Modell berechnet werden, welches eine Aussage über den aktuellen Leistungsstatus einer Person macht. Durch die Kombination dieser unterschiedlichen Methoden soll ein möglichst breites Spektrum an Hinweisen gesammelt und aggregiert werden.

⁶Nachrichten, die sofort auf dem Bildschirm erscheinen

Ein etwas anderer Ansatz haben A. Schmidt, A. Takaluoma und J. Mäntyjärvi gewählt. Um den Störeinfluss zu minimieren, haben sie ein System entwickelt welches mit Hilfe eines WAP⁷-fähigen Mobiltelefons den aktuellen Status der angerufenen Person anzeigt. Dieser Ansatz beruht darauf, dass wie beim "Face-To-Face"-Ansprechen einer Person, der Gesprächsinitiant entscheiden kann, ob er sein Gegenüber stören will. Wird beispielsweise eine Person auf seinen WAP-Telefon angerufen, erscheint die Nachricht, dass er zurzeit in einer Besprechung ist. Der Anrufer kann nun entscheiden ob er es später nochmals versuchen will oder ob seine Nachricht so wichtig ist, dass er den Anruf trotzdem fortsetzen sollte. Im zweiten Fall wird durch eine entsprechenden Eingabe der Verbindungsaufbau gestartet und das Telefon klingelt [32].

Im Experiment von J. Fogarty et al. [22] wurden Testpersonen mit einer Kamera gefilmt, wobei auch der Ton aufgezeichnet wurde. In unregelmässigen Abständen wurden die beobachteten Personen gefragt, ob sie in diesem Moment gestört werden wollen. Sie mussten dies dann auf einer Skala von 1-5 eintragen. Im zweiten Schritt wurden die Videoaufzeichnungen von unbeteiligten Personen angeschaut und sie versuchen den Störfaktor der gezeigten Situation zu ermitteln. Die Auswertung der gesammelten Daten hat dann ergeben, dass bereits mit einem einfachen Mikrofon welches erkennt ob gesprochen wird, sehr gute Vorhersagen über die Unterbrechmöglichkeit gemacht werden konnte. Um die Genauigkeit zusätzlich noch zu erhöhen, könnten einfachste Sensoren für die Erkennung von Telefongesprächen, Computermousebewegungen und Tastatureingaben hinzugefügt werden. Die Arbeit von A. Bernstein und P. Vorburger, welche einen ähnlichen Ansatz verfolgte, beobachteten bei ihrem Experiment das Verhalten bei echten Telefonanrufe, und ob sie beantwortet wurden oder nicht. Aus diesen Daten, denn Annotationen der Gespräche und multi Sektor Bewegungserkennung, wurden dann Vorhersagen auf die Unterbrechbarkeit berechnet [35].

⁷Wireless Application Protocol [31]

3

Das Experiment

Wie bereits in der Einleitung erwähnt, hat diese Arbeit zum Ziel, herauszufinden ob es möglich ist, ein mobiles System zu konstruieren, dass die Arbeitsleistung einer Person erhöhen kann. Dabei beschränken wir uns auf die kognitive Arbeitsleistung und nicht auf die physischen Fähigkeiten. Das mobile Gerät soll die Möglichkeit haben, mit Unterstützung von Context-Awareness, zu erkennen, ob die Versuchsperson im Augenblick sehr stark in ihre Arbeit vertieft ist und es aus Sicht der Arbeitseffizienz negativ auswirken würde, sie zu unterbrechen.

Dieses Kapitel erläutert die Grundlagen für die Vorbereitung und Durchführung des Experiments.

3.1 Aufbau des Experiments

Das Experiment setzt sich aus zwei Hauptbestandteilen zusammen (siehe Abbildung 3.1. Zum einen aus der Trainings- und zum andern aus der Produktivphase. In der Trainingsphase wird versucht mittels den Sensoren, die an einen mobilen Gerät angebracht sind, Vorhersagen über die Leistungsfähigkeit der Testperson zu erstellen. Im zweiten Hauptteil, der Produktivphase, wird dann erhoben, ob das konstruierte Vorhersagemodell wirklich die Leistungsfähigkeit erhöhen kann. Um die Resultate vergleichen zu können, wird der zweite Teil des Experiments an zwei Testgruppen durchgeführt. Bei der einen werden die Vorhersagemodelle angewendet und bei der zweiten nicht. Dies soll dann zeigen, ob die eine Gruppe wirklich bessere Werte erreicht als die unveränderte zweite.

Bei den Vorbereitungen des Experiments mussten drei wichtige Bestandteile vorsichtige evaluiert werden. Dies ist der psychologische Leistungstest, die eingesetzte Infrastruktur und die Probanden. Bei allen drei Teilen mussten bestimmte Einschränkungen beachtet werden.

Bei der Suche des Test musste eine Lösung gesucht werden, welche die Leistungsfähigkeit der Testperson verlässlich abbildet, andererseits muss der Test auf einem mobilen Gerät durchführbar sein.

Bei der Technik gab es auch einige Punkte, die beachtet werden müssen. Da diese Arbeit auf vorhergehende aufbaut und ähnliche Experimente bereits durchgeführt worden sind, verwendeten wir die gleichen Geräte und Software, die von D. Benz [3] eingesetzt wurden.

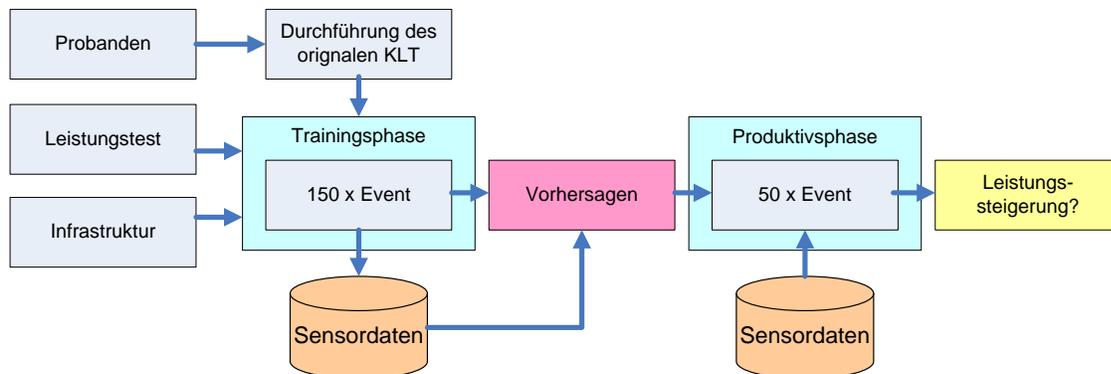


Abbildung 3.1: Phasen und Bestandteile des Experiments

Bei unserem Experiment sollten die Probanden das Testgerät den grössten Teil des Tages mit sich mitnehmen. In unregelmässigen Abständen wurden sie dann durch einen simulierten Anruf (wird im Verlauf dieser Arbeit mit Event bezeichnet) aufgefordert den Leistungstest durchzuführen. Die Resultate des Tests wie auch die Daten, welche die Sensoren aufzeichneten, wurden dann auf dem Gerät abgespeichert. Dieser Ablauf würde wiederholt, bis insgesamt 150 Events aufgezeichnet wurden. Wir haben uns für die Zahl 150 entschieden, weil uns dies für die Testteilnehmer zumutbar erschien und wir damit eine genügende Anzahl Datensätze für die Analyse erhielten.

In den nachfolgenden Unterkapitel werden nun die einzelnen Bestandteile, sowie das Experiments beschrieben. Diese sind:

- **3.2 Psychologische Leistungstests - Seite 13**
- **3.3 Technische Infrastruktur - Seite 21**
- **3.4 Die Probanden - Seite 31**
- **3.5 Durchführung des Experiments - Seite 32**

3.2 Psychologische Leistungstests

In der Psychologie gibt es unterschiedliche Vorgehensweisen, um eine Versuchsperson zu testen. Es gibt das klassische Interview, bei welchem der Testperson Fragen gestellt werden und anhand der Antworten eine Auswertung gemacht wird. Da wir in unserem Experiment mehrmals am Tag unterschiedlichste Situationen erheben wollten, konnte diese Variante für die Datenerhebung des eigentlichen Experiments nicht eingesetzt werden. Zum Schluss des Experiments wurde allerdings eine Befragung der Testkandidaten durchgeführt, welche die Form eines Interviews hatte (Siehe 4.2.3 und A.1).

Die zweite Erhebungsart, mit welcher in der Psychologie gearbeitet wird, ist die Beobachtung des Verhaltens in natürlicher oder künstlicher Umgebung. Auch diese Methode konnte bei unserem Vorgehen nicht angewendet werden, da eine Beobachtung rund um die Uhr nicht umgesetzt werden konnte. Dadurch, dass wir auch kein fest installiertes Versuchssystem einsetzten, sondern ein mobiles Gerät, hätte dies zudem nicht ohne Probleme gemacht werden können.

Als drittes psychologisches Vorgehen gibt es noch die Aufzeichnung von psychologischen Ausprägungen. Dies war dann auch die Variante, welche wir für unsere Datenerhebung verwendet haben.

Um unsere Messwerte erheben zu können, musste eine Testvariante gesucht werden, die wir beim Experiment einsetzen konnten. In der Psychologie gibt es fünf bekannte Tests die häufig angewendet werden. Dies sind Intelligenz-, Leistungs-, Neuropsychologische-, Persönlichkeits- und Verhaltenstests. Da in unserem Experiment nicht die generelle kognitive Leistungsfähigkeit der Testpersonen bestimmt werden sollte (wie dies bei einem Intelligenztest gemacht wird), sondern die Konzentrationsleistung zu einem vorgegebenen Zeitpunkt, konzentrierten wir uns auf die im nächsten Abschnitt folgenden Leistungstests [16]. Es werden dabei drei unterschiedliche Methoden der Konzentrationsmessung erläutert und es wird der optimalste für den Einsatz in Experiment evaluiert. Die Arbeit beschränkt sich auf diese drei gängigsten Tests, um möglichst gute Resultate zu erreichen. Diese sind:

- **3.2.1 Konzentrations-Leistungstest - Seite 13**
- **3.2.2 Aufmerksamkeits- und Belastungstest d2 - Seite 16**
- **3.2.3 Frankfurter Aufmerksamkeits-Inventar FAIR - Seite 17**

3.2.1 Konzentrations-Leistungstest

Der KLT¹ wurde von H. Düker und G. A. Lienert (1965) entwickelt und soll die allgemeine psychische Leistungsfähigkeit einer Testperson ermitteln. Dabei wurde versucht den Test möglichst

¹Der Test durfte nicht dem Anhang angefügt werden, da jegliche Vervielfältigung verboten ist

unabhängig von der Intelligenz zu gestalten. Der Test wurde in Hinblick auf pharmakopsychologischen Arbeiten und dem Konzept der Koordination entwickelt. Um die KLT-Aufgaben erfolgreich lösen zu können, muss diese Koordination umgesetzt werden. In der nachfolgenden Aufstellung sind die Einzelnen Komponenten, die zusammenspielen sollten, aufgelistet.

- **Auffassung:** Es müssen die Zeichen und Ziffern richtig aufgefasst werden.
- **Rechnen:** Die Rechnungen müssen richtig gerechnet werden.
- **Merken:** Die Zwischenresultate müssen kurzfristig im Kopf behalten werden.
- **Addieren, Subtrahieren:** Es muss anhand bestimmter Vorgaben entschieden werden, mit welcher Operation die Zwischenresultate behandelt werden muss.
- **Wollen:** Es muss nach der Entscheidungsfindung der richtigen Operation ein Willensentschluss gefasst werden, um die Rechnung zu vervollständigen.
- **Vorstellung:** Diese Entscheidung setzt wiederum voraus, dass sich die Testperson, die Zwischenresultate vorstellen und vergegenwärtigen kann.

Nur wenn all diese Tätigkeiten zusammenwirken, ist es möglich die KLT-Aufgaben richtig zu lösen [36].

Testaufbau

Der KLT besitzt zwei unterschiedliche Schwierigkeitsstufen. Die Version C für Grundschüler, die Version B für Jugendliche und Erwachsene. Für beide dieser Testversionen gibt es jeweils einen A- und B-Test. Diese unterscheiden sich nur in der Auswahl der Aufgaben, sind jedoch sonst identisch. So ist es möglich, dass die Resultate miteinander verglichen werden können. Für die Durchführung des Tests ist eine Bearbeitungszeit von 30 Minuten vorgesehen. Mit entsprechenden Reliabilitätseinbußen ist es auch möglich, den Test in nur 20 Minuten zu absolvieren. Der Test kann mit Personen ab dem 11. Lebensjahr durchgeführt werden [2]. Der KLT besteht aus fünf Testbögen mit je 50 zweizeiligen Rechnungsaufgaben. Bei der Testentwicklung wurden diese Aufgaben aus einem Pool von 2500 Aufgaben so ausgewählt, dass alle etwa den gleichen Schwierigkeitsgrad besitzen. Für die Auswertung der Test kann eine Lösungsschablone verwendet werden. Die Auswertung der Resultate erfolgt anhand von zwei Kriterien. Einerseits wird die Gesamtzahl gelöster Aufgaben (Leistungsrohwert $RW(L)$) erhoben und andererseits die Leistungsgüte (Fehlerrohwert $RW(F)$) des Tests. Um nun die aussagekräftige Leistungsnorm eines Probanden zu ermitteln, werden Normtafeln verwendet, die das Alter, die Schulbildung und das Geschlecht unterscheiden [9].

Der KLT ist aus Aufgaben des folgenden Typs aufgebaut (Abbildung 3.2):

Beispiel A: $8 + 9 - 2$ <input type="checkbox"/>	Beispiel B: $3 + 6 - 8$ <input type="checkbox"/>
$5 - 4 + 3$ <input type="checkbox"/>	$9 + 1 + 7$ <input type="checkbox"/>

Abbildung 3.2: Beispiel von zwei KLT-Aufgaben [9]

Zuerst müssen beide Zeilen einzeln ausgerechnet und die Resultate im Kopf behalten werden (Beispiel A: 15 bzw. 4; Beispiel B: 1 bzw. 17). Nun müssen diese Teilergebnisse nach folgenden Vorschriften weiterbehandelt werden:

- Das Untere vom Oberen abziehen, wenn es kleiner ist als dieses (wie in Beispiel A; Lösung deshalb 11).
- Das Untere zum Oberen dazuzählen, wenn es grösser ist als dieses (wie in Beispiel B; Lösung deshalb 18).

Bei der einfacheren Version C des Tests, welche für Grundschüler zu benutzen ist, muss immer das Niedrigere der beiden Resultate vom Grösseren abgezogen werden. Das Resultat ist in das dafür vorgesehene Kästchen einzutragen [2].

Reliabilität

Die Wiederholungsuntersuchung des KLT nach 8 Wochen, hat bei Oberschülern ergeben, dass die Stabilität von $r_{tt} = 0.87$ beträgt. Die Stabilität der Fehler variiert bei den Schülern entsprechend des Alters zwischen 0.74 bis 0.05. Somit kann die Fehlerzahl nicht als Leistungsgütekriterium angesehen werden [36].

Validität

Für die Beurteilung der Validität wurden aus der Gesamtheit von Schülern diejenigen ausgewählt, die 15% über oder unterhalb ihrer Altersnorm abgeschnitten hatten ($N=66$). Die tetrachorische Korrelation zwischen den Eigenschaftsbeurteilung (aktiv, passiv), die durch die Lehrer der Schüler abgegeben wurden, und der Testleistung (über-, unterdurchschnittlich) ergab eine Gültigkeitskoeffizienten von $r = 0.50$. Dies ist so zu verstehen, dass in der Regel aktive Schüler eher hohe Leistungen und passive Schüler vermehrt niedrige Leistungen im KLT erzielt hatten [9].

Fazit

Obwohl der KLT in der Praxis relativ selten eingesetzt wird, zählt er zu den besten Konzentrationstests. Er ist ein sehr verlässlicher Test, der sich hervorragend eignet, um komplexe Konzentrationsprozesse über längeren Zeitraum zu erfassen [2]. Da die Reliabilität des prozentualen

Anteils von falsch beantworteten Aufgaben zwischen $r_{tt} = 0.05$ und 0.65 liegt ist es fraglich, ob diese Kennzahl zur Bewertung der Testpersonen herangezogen werden sollte. Um die Validität des Tests zusätzlich zu stützen, empfiehlt G. Weise die Zuhilfenahme eines weiteren, andersartigen Konzentrationstests. So zum Beispiel den d2 (siehe Abschnitt 3.2.2) [36].

3.2.2 Aufmerksamkeits- und Belastungstest d2

Der d2-Test von R. Brickenkamp (1970) ist eine Weiterentwicklung des Durchstreichtests von B. Bourdon (1902). Es handelt sich hierbei um einen intelligenzunabhängigen Detail-Diskriminations-Versuch, der die visuelle Aufmerksamkeitsanspannung und die Konzentrationsfähigkeit einer Testperson prüft [36].

Testaufbau

Der d2-Test besteht aus 14 Zeilen zu je 47 d 's und p 's mit entsprechen 0, 1 oder 2 Apostroph darüber, beziehungsweise darunter (siehe Abbildung 3.3). Die Aufgabe in diesem Test ist es nun, möglichst schnell, nur diese d 's durchzustreichen die zwei Apostroph besitzen. Dabei ist es egal ob sie sich darüber, darunter oder an beiden Orten befinden. Die Testdauer beträgt etwa 5 Minuten und kann mit Personen zwischen dem 9. und 59. Lebensjahr durchgeführt werden. Für die Auswertung kann eine Schablone verwendet werden. Es wird dabei die Zahl der total bearbeiteten (GZ) sowie die Anzahl von falsch beantworteten Zeichen erhoben. Die GZ und der fehlerbereinigten Gesamtmenge (GZ-F) werden dann in die Standard- und Prozentrangnormen umgewandelt. Der Standard d2-Test wurde 1991 durch Oehlschlägel und Moosbrugger mit zusätzlichen Messwerten ergänzt, um dem instruktionswidrigen Überspringen von Testitems gerecht zu werden. Diese sind GZ-2F und Konzentrationsleistungswert, welcher die Gesamtzahl der Verwechslungsfehler repräsentiert [2].

"	"	'		"	"	'		'	
d	p	d	d	d	d	p	d	d	p
		'	'		"	"	"	'	

Abbildung 3.3: Beispiel einer d2-Aufgaben

Reliabilität

Im Mittel liegt die innere Konsistenz der Gesamtleistung bei 0.93 . Die fehlerbereinigte Leistung bei 0.91 . Diese mit Hilfe der Zeitfraktionierungsmethode erhobenen Koeffizienten sind allerdings in der Regel etwas überhöht. Bei einer Testwiederholung nach ein paar Stunden liegt der Koeffizient für die Gesamtmenge, je nach Länge des Wiederholungsintervalls, zwischen 0.71 und 0.94 . Die Stabilität für GZ-F und KL bewegen sich in der gleichen Größenordnung [2].

Validität

Es kann mit Gewissheit gesagt werden, dass eine hohe Leistung im d2-Test durch eine Testperson nur dann erzielt werden kann, wenn er seine Aufmerksamkeit intensiv und ungestört einsetzen kann. Er ist ein gutes Testmittel um die kurzfristige Aufmerksamkeitsleistung unter dem Geschwindigkeits- und Qualitätsaspekts zu messen. Die empirische Validität des d2-Tests wurde durch eine Vielzahl von Belegen bestätigt. Der Test korreliert zudem zu 0.58 mit den Resultaten aus dem KLT (siehe Abschnitt 3.2.1). Die Korrelation zu Intelligenztests ist eher gering. Völlige Unabhängigkeit ist aber nicht möglich, da die Konzentration und die Bearbeitungsgeschwindigkeit grundlegende Voraussetzungen für die Absolvierung eines Intelligenztests sind [2].

Fazit

Obwohl der d2-Test sehr ökonomischer und den Testpersonen gut zumutbar ist, widersprechen sich die Meinungen, ob er als Einzelinstrument eingesetzt werden sollte. Es ist unbestritten, dass der Test über die Zeit stabil ist. In vielen Fällen kann er als Konzentrationstest angewendet werden, der auf die kurzzeitige Anspannungsleistung abzielt. Es kann allerdings Situationen geben, in denen zusätzlich zum d2-Test noch ein anderer Konzentrationstest gemacht werden sollte. Die Kombination aus beiden wird dann ein sehr verlässliches Resultat über die Probanden abgeben [2][36].

3.2.3 Frankfurter Aufmerksamkeits-Inventar FAIR

Das Frankfurter Aufmerksamkeits-Inventar zur Messung der Konzentration wurde von H. Moosbrugger und J. Oehlschläger im Jahr 1996 entwickelt. Der Test basiert auf drei Reizdimensionen. Diese sind die Gestalt und die Punkte-Anzahl, welche als Variationen im Test vorkommen. Als dritte ist die Anordnung der Punkte, welche ausgeblendet werden müssen [2].

Testaufbau

Der Test besteht aus zwei Testbögen mit je 16 Zeilen à 20 Items. Der ganze Test ist aus vier Itemarten in je zwei Versionen zusammengestellt. Zwei dieser vier fungieren als Zielitems (siehe Abbildung 3.4). Dies können zum Beispiel ein Kreis mit zwei Punkten oder ein Quadrat mit drei Punkten sein. Bei der Testdurchführung soll der Proband möglichst schnell und ohne Fehler zwischen Ziel- und Nicht-Zielitems unterscheiden. Das ganze ist im Sinne des vollständigen Markierungsprinzips ausgelegt, wobei Nicht-Zielitems mit einer Linie unter dem Zeichen entlang geführt und bei Zielitems die Linie zackenförmig zum Zeichen hin geführt wird

		Gestalt			
		Kreis		Quadrat	
Punkte- Anzahl	3				
	2				

Abbildung 3.4: Beispiel der FAIR-Zielitems

Das FAIR kann in Gruppen oder mit Einzelpersonen durchgeführt werden, wobei die Testdauer etwas 10 Minuten dauert, wovon 6 Minuten reine Testdauer sind. Die Auswertung des Testbogens kann mit einer Schablone durchgeführt werden. Dabei werden vier unterschiedliche Kennwerte des Aufmerksamkeitsverhaltens ermittelt. Dies sind:

- **Leistungswert L:** Menge der konzentriert, bearbeiteten Items.
- **Qualitätswert Q:** Anteil, der konzentrierten Urteile in allen abgegebenen Urteilen.
- **Kontinuitätswert K:** Ausmass der kontinuierlich, aufrechterhaltenen Konzentration.
- **Markierungswert M:** Hinweise auf angemessenes Befolgen der Instruktion.

Reliabilität

Der Leistungswert L und der Kontinuitätswert K, welche für die Leistungsdiagnostik von Nöten sind, haben interne Konsistenzen (Zeitfraktionierung) zwischen 0.90 und 0.92. Bei der erneuten Testdurchgang nach 10 Minuten liegen die Retestreliabilitäten zwischen 0.85 und 0.91. Die Testwerte M und Q, welche die falsch beantworteten Fragen mit einbeziehen, werden immerhin Werte zwischen 0.65 und 0.80 erreicht.

Validität

Um die Validität des FAIR zu prüfen, kann der Vergleich mit dem d2-Test gemacht werden. Der Leistungswert L des FAIR korreliert mit dem GZ-F des d2's zu 0.50 (siehe Abschnitt 3.2.2). Diese Verfahren, bei welchem die Übereinstimmung von Ergebnissen mit einem anderen Test verglichen werden, nennt sich Konvergente Validität.

Fazit

Das FAIR ist ein objektives, reliables, valides und ökologisches Instrument welches in der Aufmerksamkeitsdiagnostik eingesetzt wird. Zudem ist das FAIR sehr verfälschungssicher.

3.2.4 Evaluation der Leistungstests

Bei der Auswahl des Tests, welcher im Experiment zur Anwendung kam, gab es einige grundlegende Einschränkungen, die beachtet werden mussten. Ursache dafür war zum einen das mobile Gerät und zum anderen, die Situation in der sich die Testperson befindet. Es musste sowohl der Test, als auch die Umsetzbarkeit auf dem Gerät bei der Suche des optimalen Tests miteinbezogen werden. Folgende wichtigen Vorgaben sollten erfüllt werden.

Ansprüche an psychologischen Test:

- **Einfachheit:** Der Test muss möglichst einfach und leicht verständlich sein.
- **Vorwissen:** Es sollte möglichst kein Vorwissen für die Bearbeitung von Nöten sein.
- **Grosser Pool:** Es müssen genügend viele Testfragen vorhanden sein.
- **Wenig monoton:** Der Test sollte so gestaltet sein, dass die Mitarbeit am Experiment nicht durch Langeweile belastet wird.
- **Sensibilität:** Der Test sollte die aktuelle Situation genügen fein abbilden können.

Technische Einschränkungen (mobiles Gerät)

- **Grafik:** Test muss auf dem kleinen Display darstellbar und leserlich sein.
- **Eingabe:** Die Eingabe am Gerät muss schnell und eindeutig sein.
- **Verarbeitung:** Die Eingaben müssen vom Gerät erkannt und gut zu verarbeiten sein.

Um nun einen passenden Test für die Umsetzung des Experiments zu finden, mussten diese Einschränkungen mit den in Frage kommenden Testeigenschaften abgestimmt werden. An dieser Stelle bietet sich eine Entscheidungsmatrix (siehe 3.1) an, mit deren Unterstützung der beste psychologische Leistungstest eruiert werden konnte.

Die Auswertung zeigte, dass der KLT bei den gegebenen Einschränkungen am besten geeignet ist (23 Punkte). Seine grössten Vorteile liegen hauptsächlich bei den technischen Einschränkungen. Weil bei diesem Test mit Zahlen und nicht mit Symbolen, wie bei den anderen beiden Tests, gearbeitet wird, ist dies ein grosser Pluspunkt für die Implementierung, wie auch für die Darstellung am kleinen Display. Durch die Vielseitigkeit der Rechnungen, wird der KLT den Probanden weniger schnell als monoton erscheinen. Der d2-Test hat hier beispielsweise weniger Variationsmöglichkeiten.

Ein eher negativer Aspekt des Konzentrations- und Leistungstest ist das Vorwissen, dass für den Test gebraucht wird. Die Testpersonen müssen die Grundrechenoperationen beherrschen und

	KLT	d2	FAIR
Einfachheit	2	3	1
Vorwissen	2	3	3
Grosser Pool	3	2	3
Wenig monoton	2	1	1
Sensibilität	3	2	2
Grafik	3	2	2
Eingabe	3	2	1
Verarbeitung	3	2	1
	23	17	14

Tabelle 3.1: Entscheidungsmatrix für Konzentrationstest (3 = beste Note)

rasch ausführen können. Dadurch, dass wir das Experiment aber nur an normal bis überdurchschnittlich gebildeten Personen durchgeführt wurde, konnte dieser Punkt vernachlässigt werden.

3.3 Technische Infrastruktur

Dieses Unterkapitel befasst sich mit den technologischen Mitteln, die für die Durchführung des Experiments gebraucht werden. Dadurch, dass bei dieser Arbeit die eingesetzte Technologie einen hohen Stellenwert hat, war es wichtig die verwendeten Komponenten sorgfältig auszuwählen. Durch die immer kleiner werdenden Geräte, mit stetig grösserer Speicher und Rechenleistung, können heutzutage Versuche umgesetzt werden, die vor ein paar Jahren noch völlig unmöglich gewesen wären. Dank dem Fortschritt in der Technik können im Bereich der Context-Awareness Untersuchungen gemacht werden, die erst durch die Verkleinerung der Sensoren und Messgeräten möglich geworden sind. In der Arbeit von D. Benz [3] wurde beschrieben welche Mittel, für die technologische Umsetzung seines Experiments am besten eingesetzt werden. Da sich diese als zuverlässig erwiesen haben, wurden am Experimentaufbau für diese Arbeit nichts, bis auf Softwareanpassungen verändert. Folgende Bestandteile werden besprochen:

- **3.3.1 Palm Tréo 750 - Seite 21**
- **3.3.1 Das Sensorboard - Seite 22**
- **3.3.1 Global Positioning System (GPS) - Seite 24**
- **3.3.2 Entwicklungsumgebung und Produkte von Drittherstellern - Seite 25**
- **3.3.2 CADT-Applikation - Seite 25**
- **3.3.2 KLT-Applikation - Seite 28**

3.3.1 Die Hardware

Dieses Unterkapitel beschäftigt sich mit den Hardwarekomponenten die für das Experiment gebraucht wurden.

Palm Tréo 750

Als Grundgerät für das Experiment haben wir den aktuell, auf dem Markt erhältlichen Tréo 750 von der Firma PalmOne². eingesetzt (siehe Abbildung 3.5). Die Wahl fiel auf dieses Gerät, da bereits in vorausgegangen Arbeiten Geräte der Firma PalmOne gesetzt wurde. Damals wurden Tréo 650, eingesetzt. Da der Tréo 650 wie auch der neue 750er über den gleichen MultiConnector Anschluss verfügen, konnten für diese Arbeit die gleichen Sensorboards verwendet werden, die bereits damals Anwendung gefunden haben [3].

²<http://euro.palm.com/de/>, Abrufdatum: 28.10.2007



Abbildung 3.5: Das Experimentiergerät - Der Tréo 750

Als Betriebssystem wird auf dem Gerät Windows Mobile ®5.0³. eingesetzt, was die Programmierung in einigen Punkten vereinfacht hat. Das Vorgängermodell setzte noch das Palm-OS 5.4.8 ein, welches beispielsweise kein echtes Multithreading eingesetzt hat [3]. Durch den 300 Mhz Samsung Hauptprozessor und die 128 MB ROM / 64 MB RAM, verfügt das Gerät über genügen Rechen- und Speicherleistung, um unterschiedlichste Anwendungen auf dem Gerät einsetzen zu können. Ein weiterer Vorteil des Tréos ist das grosse, mit einer Touchscreen-Funktion ausgerüstete Display. Seine 4,8 cm x 4,8 cm Seitenlänge bieten sich sehr gut an, um bequem Dateieingaben zu tätigen oder komplexe Darstellungen anzuzeigen. Durch den Touchscreen, welcher mit einem Stift oder sogar mit dem Fingern bedient werden kann, können Eingaben bequem auf dem Display erfolgen und es muss nicht zwingend die eingebaute Gerätetastatur benutzt werden. Für die persistente Speicherung verfügt der Tréo neben dem internen Speicher über ein MiniSD Karteneinschub. Für das Experiment wurde die maximal unterstützte Speichergrösse von 2 GB gewählt [21].

Das Sensorboard

Bei dieser Arbeit spielt das von Peter Vorburger⁴ entwickelten Sensorboard (siehe Abbildung 3.6) eine zentrale Rolle. Es wurde mit einer Vielzahl von Sensoren ausgestattet die verschiedenste Messwerte liefern können. Um diese Daten vom Sensorboard zu übertragen, verfügt es über ein MultiConnector. Dieses Verbindungsstück kann dann in unserem Fall an verschiedensten Produkten der Firma PalmOne angebracht werden, womit man sich eine Vielzahl von möglichen Einsatzszenarien ergeben. Da das Sensorboard den nötigen Strom direkt aus dem Akku des Tréos

³<http://www.microsoft.com/germany/windowmobile/default.aspx>, Abrufdatum: 28.10.2007

⁴Peter Vorburger, Doktorand Institut für Informatik, Universität Zürich, <http://www.ifi.uzh.ch/ddis/people/vorburger/>

bezieht, ist nur eine Stromquelle für die Kombination Tréo/Sensorboard nötig. Dies ist ein Vorteil, da so keine zusätzliche Stromversorgung für die Sensoren von Nöten ist. Somit kann der Ablauf des Experiments auch nicht damit belastet werden, dass die Sensordaten durch einen Mangel an Energie nicht aufgezeichnet werden. Entweder haben beide Komponenten genügend Strom oder beide unterbrechen gleichzeitig die Arbeit.

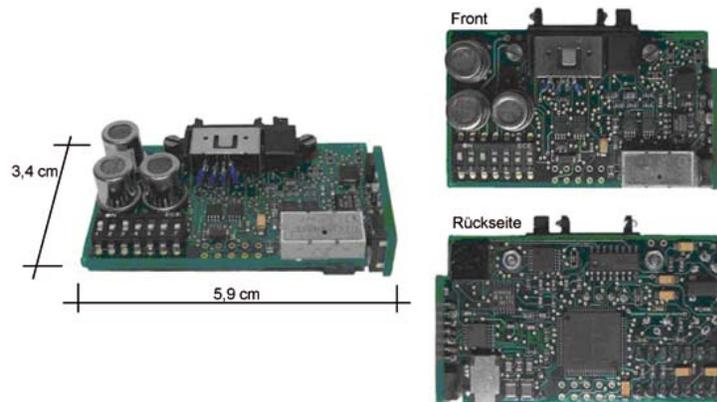


Abbildung 3.6: Das Sensorboard

Das Sensorboard ist mit folgenden Sensoren ausgestattet:

1. **Temperatur**
2. **Accelerometer x-Achsen**
3. **Accelerometer y-Achsen**
4. **Accelerometer z-Achsen**
5. **Gyroskop x-Achsen**
6. **Gyroskop y-Achsen**
7. **Gyroskop z-Achsen**
8. **Gas Sensor I (CO)⁵**
9. **Gas Sensor II/III (Brennbare Gase/Alkohol)**
10. **Magnetfeldsensor x-Achsen**
11. **Magnetfeldsensor y-Achsen**

⁵Kohlenstoffmonoxid

12. Magnetfeldsensor z-Achsen

Die Sensoren die auf dem Sensorboard angebracht sind besitzen analoge Schnittstellen. Damit nun über die digitale Schnittstelle mit dem Tréo kommuniziert werden kann, werden die Daten im Analog-Digital-Wandler umgewandelt. Dieser liefert dann einen Zahlenwert zwischen 0 und 4096 (12-bit), welcher über die MultiConnector-Schnittstelle gesendet und so im Gerät weiterverarbeitet werden kann.

Die Daten, die das Sensorboard dem Verarbeitungsgerät (in unserem Fall der Tréo 750) sendet, sind durch Komma getrennte, hexadezimale Werte. Diese werden in Zeilen zu jeweils 12 Werten übermittelt. Die Übertragung erfolgt durchschnittlich etwa 20 Mal in der Sekunde. Die Abbildung 3.7 zeigt ein kurzer beispielhafter Auszug der Daten die das Sensorboard übermittelt.

```

...
08.09.2007 16:18:59-FFA,E99,EA8,00C,741,9A2,92B,245,249,001,005,001-47°38'59.1125''N 009°11'10.9225''E
08.09.2007 16:18:59-FFA,E9C,E98,009,73E,9A1,925,245,238,001,000,005-47°38'59.1125''N 009°11'10.9225''E
08.09.2007 16:18:59-FFA,E92,EA1,006,741,9A2,928,244,246,009,008,003-47°38'59.1125''N 009°11'10.9225''E
08.09.2007 16:18:59-FFA,EA3,E9B,00A,738,9A3,91F,248,23F,009,00F,002-47°38'59.1125''N 009°11'10.9225''E
...

```

Abbildung 3.7: Datensatz des Sensorboards

Global Positioning System (GPS)

Der Versuchsaufbau, auf dem diese Arbeit aufbaut [3], hatte vorgesehen, dass ein GPS-Gerät für die Positionsermittlung des Geräts verwendet wird. Aus diesem Grund wurde auch bei dieser Arbeit ein GPS-Modul verwendet, um den Aufenthaltsort einer Versuchsperson zu ermitteln.

Das GPS-Gerät, heisst Altina GBT-709 und wurde von der Firma Union Rich Development entwickelt⁶. Es verfügt über eine Bluetooth- und USB-Schnittstelle über welche mit anderen Geräten kommunizieren werden kann. In unserem Fall haben wir uns für die Bluetooth Variante entschieden. Der Grund dafür liegt hauptsächlich darin, dass der Tréo über keine USB-Schnittstelle verfügt (ausser für die Kommunikation mit einem Computer). Da er allerdings ein Bluetoothmodul integriert hat, bietet sich dieses auch gleich an, um darüber zu kommunizieren. Das GBT-709 verfügt über das Serial Port Profile (SPP), welches eine serielle Schnittstelle emuliert und mit Bluetooth überträgt. Mit diesen seriellen im Verarbeitungsgerät ankommenden Datensätzen kann dann weitergearbeitet werden. Ein solcher Datensatz wird im NMEA (National Marine Electronic Association) Protokoll gesendet, wobei das GGA (Global Positioning System Fix Data, Time, Position and fix related data for a GPS receiver) Format verwendet wird.

Der im GPS-Empfänger eingebaute 3,7 Volt 850 mAh Akku soll nach Herstellerangaben etwa für 12 Stunden Betrieb reichen. In der Praxis wird allerdings nur ein Wert von ca. 8h erreicht. Da

⁶<http://www.altina.com.tw>, Abrufdatum: 28.10.2007

das GPS eigentlich für den Betrieb im Freien entwickelt wurde, ist der Empfang zu den Satelliten in Gebäuden stark eingeschränkt oder sogar unmöglich. Aus diesem Grund ist es erstaunlich, wie gut der GBT-709 auch in Gebäuden arbeitet. Es hat sich gezeigt, dass wenn einmal eine Verbindung zu den Satelliten aufgebaut wurde, auch ein Betrieb in der Nähe der Fenster möglich ist. Somit können teilweise auch in den Gebäuden die Positionierungsdaten der Experimentteilnehmer erfasst und aufgezeichnet werden.

3.3.2 Die Software

Das in dieser Arbeit entwickelte Experiment wurde in dieser Form noch niemals durchgeführt. Aus diesem Grund gibt es keine bereits existierende Software die dafür hätte eingesetzt werden können. In der Arbeit von D. Benz [3] wurde deshalb eine Applikation entwickelt, die Sensordaten aus dem bereits erwähnten Sensorboard (siehe Abschnitt 3.3.1) aufzeichnet. In diesem Unterkapitel wird nun auf diese Applikation und dessen Programmierung eingegangen. Gleichzeitig werden die Anpassungen und Ergänzungen aufgezeigt, die im Zuge dieser Arbeit gemacht werden mussten.

Entwicklungsumgebung und Produkte von Drittherstellern

Die Grundapplikation welche für das Experiment eingesetzt und angepasst wurde, ist in C# für Windows Mobile 5.0 programmiert. Für die Entwicklung ist Microsoft Visual Studio 2005⁷ ausgewählt worden. Zusätzlich wurden bei der Umsetzung Frameworks und Libraries eingesetzt, um bestimmte Funktionalitäten in der Applikation einzubauen. In der Tabelle 3.2 sind die eingesetzten Produkte ersichtlich.

Produkt	Einsatz	Link
OpenNETCF	Diverse Libraries	OpenNETCF.org ⁸
GPS.NET API	Library für objektorientierten Zugriff auf GPS	GeoFrameworks ⁹
MortSaver	Verhindert das aktivieren des Standby-Betriebs	Mirko Schenk ¹⁰

Tabelle 3.2: Verwendete Software von Drittherstellern

CADT-Applikation

Die CADT-Applikation¹¹ diente als Grundapplikation des durchgeführten Experiments. Um den geänderten Anforderungen gerecht zu werden mussten einige Anpassungen am Code vorgenommen werden. In diesem Abschnitt wird nur die CADT-Applikation vor der Anpassung beschrieben, danach folgen die Änderungen die gemacht wurden (siehe Tabelle 3.3).

⁷<http://msdn2.microsoft.com/de-de/vstudio/default.aspx>, Abrufdatum: 28.10.2007

⁸<http://www.opennetcf.com/>, Abrufdatum: 28.10.2007

⁹<http://www.geoframeworks.com>, Abrufdatum: 28.10.2007

¹⁰<http://www.sto-helit.de>, Abrufdatum: 28.10.2007

¹¹CADT steht für Context Awareness Disturbing Test

Funktionen der CADT-Applikation Die Hauptaufgabe der CADT besteht darin Daten aufzuzeichnen und abzuspeichern. Diese Daten setzen sich aus den unter 3.3.1 erwähnten Messdaten der Sensoren, den GPS-Informationen und den Audioaufzeichnungen des Geräts zusammen. Weiter verfügt die Applikation über eine Funktion, mit der Telefonanrufe simuliert werden können. Durch die Simulation kann beeinflusst werden, in welchen Abständen die Anrufe stattfinden, was bei Echten, nicht künstlich erzeugten, nicht gehen würde. Dadurch wird der Sollwert der Anzahl Anrufe, die für eine wissenschaftliche Erhebung von Nöten sind, meist schneller erreicht.

Die Applikation hat den Zweck, Sensordaten jeweils bis zu einem simulierten Telefonanruf aufzuzeichnen und dann auf die Speicherkarte abzuspeichern. Wie lange aufgezeichnet wird, hängt gleich wie die Anzahl Anrufe, von den Einstellungen ab, die frei gewählt werden können. Bei jedem Anruf wird vor dem Annehmen des Anrufs abgefragt, wie stark dieser Anruf im Moment stört (im weiteren Verlauf dieser Arbeit auch Unterbrechunglevel genannt). Nach der Eingabe des Störlevels, auf einer Skala von 1 bis 5, kann der Anruf angenommen werden.

Die Grund-Applikation, die im Zuge dieser Arbeit angepasst wurde, hatte die Aufgabe herauszufinden, ob es möglich ist, anhand der erhobenen Sensordaten, die Lokalität sowie die Aktivität der Versuchsperson zu ermitteln [3]. Aus diesem Grund verfügt das Programm über eine Funktionalität, die nach dem Auflegen eines Anrufs den Benutzer fragt, wo er sich gerade befindet und welcher Aktivität er nachgeht. Diese Eingaben werden dann zusammen mit den gespeicherten Sensordaten in einem Datenpaket abgelegt.

Anpassungen der CADT Die Modifikationen, die im Zuge dieser Arbeit an der Applikation vorgenommen wurden, sind mehrheitlich nur Anpassungen um dem geänderten Experimentaufbau gerecht zu werden. Neben Änderungen an den Funktionen der Applikation, wurden auch grafische Modifikationen vorgenommen. Da die Software nicht mehr nur Prototyp-Charakter hatte und mit Probanden durchgeführt wurde, wurde das GUI¹² ein wenig ansprechend gestaltet (siehe Abbildung 3.8).



Abbildung 3.8: GUI der CADT-Applikation

¹²Graphical User Interface

In der nachfolgenden Tabelle 3.3 sind die Programm-Klassen mit den Anpassungen ersichtlich.

Klasse	Anpassungen
Ganze Applikation	<ul style="list-style-type: none"> • Übersetzung der englischen Ausdrücke (nur für Benutzer sichtbar) • Optimierung der Bedienung und Benutzerfreundlichkeit • Minimierung von möglichen Fehlverhalten der Benutzer
MainForm	<ul style="list-style-type: none"> • Neues GUI: <ul style="list-style-type: none"> - keine Sensordaten auf dem Display - Aktivitätsanzeige - Exit-Button für Unterbruch • Zufallsfunktionalität für die Anrufsimulation • Entfernen der Aktivitäts- und Lokalisierungsmethoden • Anpassungen für simulierten Anruf • Anpassungen um Lebensdauer des Akkus zu erhöhen
Simulator	<ul style="list-style-type: none"> • Anpassungen für simulierten Anruf
IncomingCallForm	<ul style="list-style-type: none"> • Umgestaltung GUI • Startprozedur für KLT-Applikation
Simulator	<ul style="list-style-type: none"> • Debugging Audioausgabe
ActivityForm	<ul style="list-style-type: none"> • Deaktivierung dieser Funktionalität
LocationForm	<ul style="list-style-type: none"> • Deaktivierung dieser Funktionalität
AudioRecorder	<ul style="list-style-type: none"> • Änderung Dateiformat um Wiedergabe auf Tréo zu verhindern
TalkingForm	<ul style="list-style-type: none"> • Deaktivierung dieser Funktionalität

Tabelle 3.3: Anpassungen CADT-Code

Wie bereits erläutert, wurden bei der Entwicklung der CADT-Applikation Frameworks und Libraries von Drittanbietern verwendet. Die Library für die Funktionalität des GPS-Empfängers wurden von der Firma GeoFrameworks [20] bezogen. Da diese kostenpflichtig sind, haben wir eine Academic Licence für 6 Monate beantragt. Der aktuelle für diese Arbeit benutzte Produkteschlüssel ist noch bis Freitag 29. Februar 2008 gültig. Eine Verlängerung kann für akademische Zwecke jederzeit bei GeoFrameworks beantragt werden. Der Produkteschlüssel befindet sich im Applikationsordner Properties in der Datei AssemblyInfo.cs.

KLT-Applikation

Um nun die Leistungsfähigkeit der Versuchspersonen messen zu können, musste eine Applikation entwickelt werden, die dies ermöglicht. Wie bereits erwähnt wurde dazu der Konzentrations-Leistungstest ausgewählt (siehe Abschnitt 3.2.4). Bei der Umsetzung musste darauf geachtet werden, dass das Testprogramm für den Tréo möglichst nahe am ursprünglichen Test liegt, um die Aussagekraft der Resultate nicht zu verfälschen. Da bereits bei der Auswahl des eingesetzten Leistungstest, Überlegungen zur grafischen Umsetzung gemacht worden sind, konnte nahezu die originale Darstellung verwendet werden. Einzig die Listenform der Aufgaben musste auf Einzeldarstellung geändert werden, da der begrenzte Raum auf dem Display dies nicht zuließ. Die Abbildung 3.9 zeigt wie die Grafik für den KLT umgesetzt worden ist.

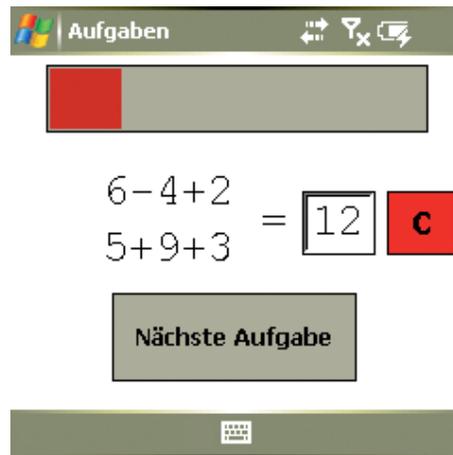


Abbildung 3.9: Darstellung des KLT auf dem Display des Treo 750

Da der KLT nicht bei jeder Datenmessung im ganzen Umfang durchgeführt werden kann, wurde er auf jeweils 10 Rechungsaufgaben pro Erhebung gekürzt. Entsprechend dazu musste auch die Bearbeitungszeit von 30 Minuten pro Leistungsmessung auf 72 Sekunden gekürzt werden. Um den Testpersonen den Leistungsdruck trotz der kürzeren Bearbeitungszeit zu vermitteln, wurde auf dem GUI ein Fortschrittsbalken angebracht, der die verbleibende Zeit bis zum Ende anzeigt. Eigenversuche haben gezeigt, dass durch dieses kleine, visuelle Hilfsmittel erstaunlich gut das Gefühl von Stress angeregt und die Leistungsbereitschaft angespornt wurde. Um sicherzustellen, dass der Testteilnehmer genau weiss wie der KLT absolviert werden sollte, wird jedes Mal beim Starten der Applikation eine kurze Anweisung mit Beispielen angezeigt. Diese kann dann gelesen oder direkt mit dem lösen des Test begonnen werden. Um sicherzustellen, dass der Test möglichst die aktuelle Leistungssituation abbildet, sind Zeitbegrenzungen einprogrammiert, die verhindern sollen, dass die Testperson sich Zeit lassen kann und sie sich der Rechensituation anpasst. Gegebenenfalls würde nach 60 Sekunden abgebrochen werden. Währenddem die Rechungsaufgaben gelöst werden, laufen im Hintergrund der KLT-Applikation unterschiedlich Prozesse ab. Der wichtigste ist die Erhebung und Abspeicherung der Testleistung. Für jedes Event

wird ein Datensatz erstellt mit den Kennzahlen, die in der untenstehenden Auflistung ersichtlich sind. Bei unserer Umsetzung des KLT werden zusätzlich zu den notwendigen noch weiterführende Zeitwerte erhoben, die Auskunft über die Leistung der Testperson geben sollen und in die Auswertung mit einfließen.

Ein Datensatz für ein KLT-Event ist aus folgenden Werten zusammengestellt:

- **Dauer in Sek für Instruktion/Testbeginn**
- **Datum mit Uhrzeit des Starts**
- Für jede einzelne Rechnungsaufgabe:
 - **Nummer des Rechnungspaares**
 - **Resultat das eingegeben wurde**
 - **Resultat das richtig ist**
 - **Bewertung des Resultats**
 - **Zeit, die für Berechnung gebraucht wird**

Wird bei der Instruktionseite, die zu Beginn des Tests angezeigt wird, zu lange gewartet, wird nach 60 Sekunden abgebrochen, wobei die Applikation aber trotzdem einen Datensatz erstellt, in welchem dieser Umstand festgehalten ist.

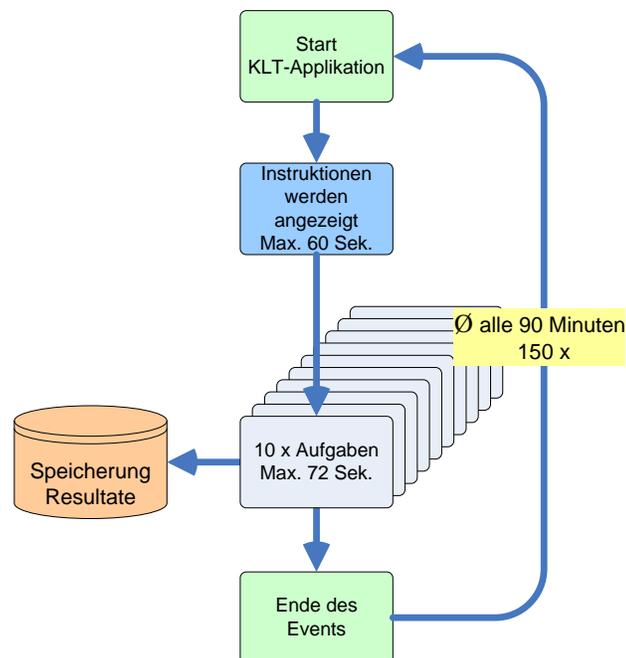


Abbildung 3.10: Ablauf der KLT-Applikation

Nach Ablauf der 72 Sekunden, die der Test maximal dauert oder nach Bearbeitung aller 10 maximal möglicher Aufgaben, werden alle erhobenen Werte auf dem internen Gerätespeicher abgespeichert und zur Sicherheit noch auf der MiniSD-Speicherkarte kopiert. Die Applikation wird dann vollständig geschlossen und beim nächsten Event wieder aus der CADT-Applikation heraus neu gestartet. Abbildung 3.10 zeigt den Ablauf der KLT-Applikation.

3.4 Die Probanden

Für die Durchführung des Experiments brauchte es mehrere Probanden. Bevor diese gefunden werden konnten, mussten die allfälligen Einschränkungen identifiziert werden, die das Resultat in irgendeiner Art negativ beeinflussen könnten. Aus der Sicht des KLT gab es zwei Vorgaben die eingehalten werden mussten. Zum einen mussten die Testteilnehmer die Addition und Subtraktion von zweistelligen Zahlen gut beherrschen. Zum anderen war es von Vorteil, wenn alle Testpersonen ein ähnlichen Bildungsniveaus¹³ aufwiesen. Dies ermöglicht es, dass die Resultate miteinander verglichen werden konnten, da nach H. Dücker die Resultate entsprechend der ausübenden Berufe variieren. Dadurch, dass der KLT nur eine sehr grobe Unterscheidung beim den Alterstufen der Versuchspersonen macht¹⁴, und wir nur Personen gesucht haben, die in einer Büroumgebung arbeiten, konnten Schüler und Jugendliche ausgeschlossen werden. Ein weiteres Kriterium für die Wahl der Experimentteilnehmer/innen war der Umgang mit dem mobilen Testgerät. Da Technik im generellen nicht ganz ohne Tücken auskommt, brauchten wir Personen, die ein gewisses technisches Wissen mitbrachten. Weil das Experiment etwa einen Monat pro Versuchsperson dauerte, und sie dabei jeden Tag das Gerät auf sich tragen mussten, konnte es Situationen geben, in denen, eine in technischer Hinsicht unerfahrene Person, sich nicht zu helfen weiss.

Während der Suche nach Testpersonen kam es vor, dass Leute gefunden wurden, die gerne am Experiment mitmachen wollten, dies aber von seitens ihres Arbeitgebers nicht geduldet worden wäre. Sie argumentierten, die Arbeitnehmer würden durch ein solches Experiment von der realen Arbeit abgelenkt und die Event-Benachrichtigung des Testgeräts würde andere Personen im Raum stark stören. Aus Sicht der Experimentvorbereitung war diese Aussage natürlich negativ, da wir auf der Suche nach genau solchen Testpersonen waren, die bei der Arbeit durch verschiedenste Kommunikationsmittel unterbrochen werden. Diese Aussage zeigt aber eindrücklich, dass das Thema Unterbrechung und Arbeitsleistung die Arbeitgeber beschäftigt und sie nicht gewillt sind, Ablenkungen ihrer Mitarbeiter ohne weiteres zu akzeptieren.

Bei anderen potenziellen Testkandidaten, wurde zudem oftmals die Dauer der Erhebung von mindestens 20 Tagen, als viel zu lange bewertet. Es ist aber verständlich, dass nicht jede Person bereit ist, über 3 Wochen hinweg für mindestens 10 Stunden am Tag den Tréo mit sich zu nehmen und im Durchschnitt alles 1,5 Stunden eine Rechnungsaufgabe zu lösen. Wir haben uns deshalb darauf geeinigt, dass die Versuchspersonen für dieses Experiment finanziell entschädigt werden sollten. Wir erhofften uns daraus, dass die Resultate durch die Bezahlung der Testteilnehmer an Qualität zulegen. Es sollte das Pflichtgefühl der Probanden angeregt werden, das Gerät möglichst oft mit sich zu nehmen und so den Versuchsablauf möglichst schnell abzuschliessen. Dadurch, dass das Gerät öfters mitgenommen wird, konnten zudem mehr unterschiedliche Situationen erfasst werden und die Anzahl an nicht beantworteten Events wurde minimiert.

¹³Die Einteilung im KLT ist: Ländliche Berufe / Arbeiter, Facharbeiter, Handwerker / Büro-, Bank-, Kaufmann / Freie Berufe / Studenten [9]

¹⁴Es wird nur zwischen Jugendlichen und Erwachsenen unterschieden. Für Schüler gibt es einen spezialisierten Test (siehe Abschnitt 3.2.1) [9]

Da wir für die Durchführung des Experiments nur eine begrenzte Anzahl von mobilen Geräten zur Verfügung hatten, wurden fünf vielversprechende Personen ausgewählt die uns am besten geeignet schienen. In der Tabelle 3.4 sind verschiedene Angaben zu den Probanden die am Experiment teilgenommen haben abgedruckt. Es wurden leider keine Teilnehmer gefunden, die in der Privatwirtschaft arbeiten und sich am Experiment beteiligen durften. Wir konnten deshalb nur Angehörige der Universität Zürich für unsere Forschungsarbeit einsetzen. Da alle Teilnehmer regelmässig Büro- und Schreivarbeiten erledigten, wird glücklicherweise die Gleichartigkeit zur "normalen" Bürosituationen erfüllt.

INDEX	GESCHLECHT	ALTER	TECHNISCHES WISSEN
#1	M	26 Jahre	Hoch
#2	M	29 Jahre	Hoch
#3	M	27 Jahre	Hoch
#4	M	26 Jahre	Hoch
#5	M	35 Jahre	Hoch

Tabelle 3.4: Angaben zu den Probanden

Die eingesetzte Technik war als Prototyp zu verstehen. Dadurch konnte es vorkommen, dass die Testteilnehmer mit Software- und Hardwareproblemen konfrontiert wurden. Aus diesem Grund ist im Vorfeld festgelegt worden, dass alle Teilnehmer während des Tests persönlich betreut werden. Diese Betreuung konnte optimal durchgeführt werden, da alles Beteiligten regelmässig an der Universität Zürich anzutreffen waren. Bedienungsprobleme oder kleine Fehler im Code, konnten so rasch und sehr kurzfristig behoben werden.

3.5 Durchführung des Experiments

Der Ablauf des eigentlichen Experiments ist in vier Schritte aufgeteilt, welche im folgenden Unterkapitel erläutert werden.

3.5.1 KLT-Eingewöhnung und Abgabe der Geräte

Ein Selbstversuch hatte gezeigt, dass die Berechnung von KLT-Aufgaben (siehe 3.2.1) nach einer kurzen Zeit schneller und einfacher ablaufen, als bei den ersten Durchgängen. Um diesem klassischen Lernkurveneffekt [11] entgegenzuwirken, haben wir vor dem Beginn des Experiments einen original KLT mit den Probanden durchgeführt. Obwohl die Endresultate (siehe 4.2) trotzdem einen leichten Lerneffekt aufzeigten, wurde dieser einführende KLT von den Probanden, als gute Vorbereitung angesehen. Dies zeigte, dass die Eingewöhnung in den Ablauf der KLT-Aufgaben erfolgreich war. Gleichzeitig konnte damit eine Verständniskontrolle durchgeführt werden, die gezeigt hat, ob die Rechnungen auch richtig angewendet worden sind.

Der Test wurde auch durchgeführt, um die Reliabilität (siehe Abschnitt 5.1) des von uns angepas-

ten KLT zu garantieren. Es zeigte sich, dass die Aufteilung in kleine Testeinheiten die Endaussage nicht veränderte.

Die Resultate des einführenden KLT wurden den Testteilnehmern vor Abschluss des ganzen Experiments nicht mitgeteilt. Damit sollte vermieden werden, dass sie eine persönliche Leistungseinstufung machen können und dies möglicherweise Auswirkungen auf die Leistung im eigentlichen Experiment hat.

Beim Beginn des Experiments hatten alle Beteiligten einen Tréo mit Sensorboard, eine Dockingstation für das Aufladen des zweiten Akkus, ein GPS-Empfänger und eine Benutzerhandbuch bekommen. Das Benutzerhandbuch (siehe Anhang A) enthält eine kurze Einführung in den Ablauf des Experiments und Erklärungen zu den eingesetzten Geräten wie auch zum KLT. Weiter beinhaltet es eine Anleitung wie mögliche Fehler vermieden und gegebenenfalls gelöst werden können. Durch die Übergabe der Geräte startet das Experiment und die Sensordaten zeichneten ab diesem Zeitpunkt auf.

3.5.2 Der Testbetrieb

Während der laufenden Tests, wurde der Versuchsteilnehmer gebeten, das Gerät jeden Morgen mit einer über Nacht aufgeladenen Batterie auszurüsten und einzuschalten. Die Software arbeitet dann den ganzen Tag hindurch und zeichnete jeweils 10 Minuten vor einem Event die Sensordaten auf. Die Audioaufzeichnungen wurden aus Gründen der Speicherkapazität¹⁵ jeweils nur 3 Minuten vor einem Event gestartet. Um der Versuchsperson anzuzeigen, dass ein Event stattfindet, war jeweils aus dem Tréo ein Klingelton zu hören und er vibriert mit Hilfe der eingebauten Vibrationsfunktion. Der Benutzer hatte dann 20 Sekunden Zeit um den simulierten Telefonanruf anzunehmen. Die Annahme erfolgte dadurch, dass auf dem Display des Tréos der aktuelle Störlevel ausgewählt wurde (siehe Abbildung 3.11). Nach der Eingabe dieses Wertes öffnete sich die KLT-Applikation und es war eine Rechenanleitung zu sehen (siehe Abbildung 3.12). Die Versuchsperson hatte dann 60 Sekunden Zeit um sie durchzulesen oder sie hatten sofort damit begonnen die Rechungsaufgaben zu lösen. Insgesamt standen 72 Sekunden zur Verfügung um die Aufgaben entsprechend der Regeln (siehe 3.2.1) zu berechnen. Sobald die Zeit abgelaufen war, oder die maximale Anzahl von 10 Rechnungen hintereinander gelöst wurden, brach das Programm ab und es dauerte erneut zwischen 30 bis 120 Minuten bis ein nächster Event angezeigt wurde. Falls ein Event nicht beantwortet wurde, erstellte das System trotzdem einen Datensatz mit den Sensordaten und markierte sie als verpasster Event. Wegen des hohen Stromkonsums der Sensoren, war es erforderlich, dass etwa alle 6 Stunden der Akku ausgetauscht werden musste. Da jede Versuchsperson über mindestens zwei Akkus verfügte, stellte dies allerdings kein Problem dar. Am Ende jeden Tages, wurde das Experiment unterbrochen und beide Akkus müssten vollständig aufgeladen werden, damit am darauffolgenden Tag das Experiment weitergehen konnte.

¹⁵Pro Event werden etwa 14 MB aufgezeichnet und abgespeichert - Dies ergibt mindestens 2,1 GB aggregierte Daten pro Proband

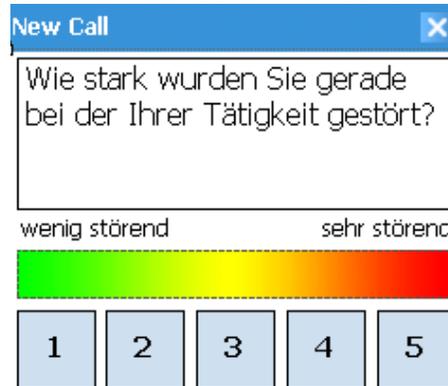


Abbildung 3.11: Auswahl des Unterbrechungslevel auf dem Tréo

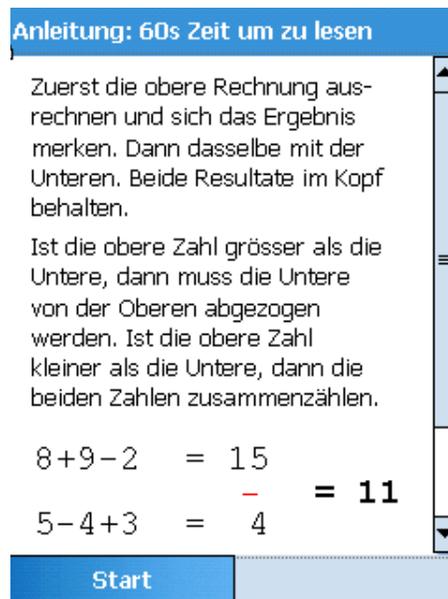


Abbildung 3.12: Darstellung der KLT-Anleitung auf dem Tréo

3.5.3 Datenübertragung

Für eine repräsentative Auswertung der Resultate wurden von jedem Probanden 150 beantwortete Events benötigt. Ein Event galt als beantwortet, sobald das Störlevel auf dem Gerät ausgewählt worden ist. Da hin und wieder Events durch die Teilnehmer verpasst worden sind, da die Person beispielsweise die Signalisierung des Geräts nicht bemerkten, wurden mehr einzelne Datenpakete aufgezeichnet, als nur die 150 Angestrebten. Im Durchschnitt war der Anteil an verpassten

Events allerdings unter 15%¹⁶. Dadurch, dass die maximale Speicherleistung auf dem Tréo nur 2 GB betrug, reichte sie nicht aus, um alle Events darauf abzulegen. Aus diesem Grund musste während des laufenden Experiments das Gerät eingezogen und die gespeicherten Daten auf ein anderes Medium verschoben werden. Dieser Vorgang dauerte jeweils nur ein paar Minuten und der Experimentablauf wurde dabei.

3.5.4 Abschluss Trainingsphase

Sobald der Sollwert von 150 beantworteten Events erreicht worden war, erschien auf dem Display des Geräts eine Nachricht, dass der Test erfolgreich abgeschlossen war. Die Testteilnehmer wurden zudem aufgefordert das Gerät dem Testverantwortlichen zurückzugeben, womit der erste Teil des Experiments abgeschlossen war.

3.5.5 Produktivphase

Das ursprüngliche Vorgehen beim Experiment mit einer Trainingsphase und anschließender Produktivphase, musste an dieser Stelle abgeändert werden. Bei der Analyse der Vorhersagen (siehe Kapitel 4), was die Verbindung dieser zwei Phasen darstellt, wurde erkannt, dass die Vorhersagen sehr ungenau waren. Da diese schlechte Werte keinen Nutzen in Punkto Leistungsfähigkeitserhöhung brachten, wurde der zweite Teil des Experiments, die Produktivphase, abgebrochen. Weitere Auführungen siehe Kapitel 5.

¹⁶Verpasste Events (pro Proband): #1: 21; #2: 12; #3: 15; #4: 17; #5: 45; => 110:5 = 22 $\hat{=}$ 14.6%

4

Auswertung der Daten und Resultate

In diesem Kapitel werden die Auswertung der Datensätze, die im Experiment gesammelt worden sind behandelt und die Resultate analysiert.

4.1 Die Aufbereitung der Daten

Die gesammelten Daten mussten alle individuell aufbereitet werden, damit sie für die Analyse in einheitlicher Form sind. In den nachfolgenden Abschnitten wird dieser Prozess erläutert.

4.1.1 Die Aufbereitung der Sensordaten

Die Datensätze, welche während des Experiments auf dem Gerät gespeichert worden sind und aus Sensorboard-, Audio-, GPS-, Unterbrechungslevel und Leistungstestdaten zusammengesetzt sind, wurden alle unformatiert auf dem Gerätespeicher abgelegt. Nach der Übertragung auf einen Desktop-PC mussten sie zuerst aufbereitet werden, bevor sie für die Analyse verwendet werden konnten. Um aus den Rohdaten (Sensorboard, GPS, Unterbrechungslevel und Audioaufzeichnungen) für die Analyse verwendbare Datensätze zu erzeugen, haben wir Matlab¹ verwendet. Beim Processing wurden aus den Sensordaten jeweils die Mittelwerte und die Standardabweichungen für jeden einzelnen Sensor ermittelt und abgespeichert. Bei der Auswertung der Audio-dateien wurden zusätzlich pro Sekunde 17 verschiedene Audio-Features berechnet und daraus jeweils auch der Mittelwert und die Standardabweichung abgespeichert. Die eingesetzten Audio-Features waren [3]:

- Spectral center of gravity

¹Matlab Version 7.5 (R2007b), The MathWorks, Inc. www.mathworks.com

- Temporal fluctuations of spectral center of gravity
- Common onsets across frequency bands
- Histogram width
- Variance
- Mean level fluctuation strength
- Zero crossing rate
- Cepstral coefficients 1-10

Eine ausführliche Beschreibung und Erklärungen zu diesen unterschiedlichen Features gibt es in der Masterarbeit von P. Egger [12]. Alle aufgearbeiteten Werte wurden dann mit Matlab in ein für die weitere Bearbeitung kompatibles Format (.arff) zusammengefasst und abgespeichert. Für jede einzelne Testperson wurden 10 Datensätze erstellt, die sich dadurch unterschieden, dass der betrachtete Zeitausschnitt vor einem Event verschieden lange gewählt ist. Wir haben uns dabei auf 5, 10, 30, 45, 60, 90, 120, 180, 240 und 300 Sekunden Aufzeichnungsdauer vor einem Event beschränkt.

4.1.2 Aufbereitung der KLT-Resultate

Neben den Daten der Sensoren mussten auch die Rechnungsleistungen, der einzelnen Konzentrations- und Leistungstests aufbereitet werden. Wie unter 3.3.2 beschrieben wurde für jeden Event ein Datensatz mit Angaben zu den Leistungen der Aufgaben angelegt und auf dem Tréo abgespeichert. Da wir den KLT nicht in seiner Urform anwenden konnten und ihn an unsere Ansprüche angepasst haben, musste eine sinnvolle Auswertungsmethode gefunden werden, da die Analyse mit der originalen Methode zu ungenau gewesen wäre. Sie hätte vorgesehen, dass nur eine Einteilung in unter-, über- und durchschnittliche Leistung im Vergleich zu einer genormten Testgruppe gemacht werden würde. Um eine möglichst umfassende Darstellung der Leistung der Testteilnehmer zu erhalten, haben wir 10 unterschiedliche Auswertungsmethoden für die Rechnungsresultate entwickelt, mit welchen die einzelnen Leistungen bei den Events berechnet wurden. Diese 10 unterschiedlichen Kennzahlen werden in dieser Arbeit mit **KLT1** bis **KLT10** bezeichnet. Die Berechnung der KLT-Werte wurden nach folgenden Regeln erstellt:

- **KLT1:**

$$\sum_1^{10} \frac{(falsche = 0 \vee richtig = 1)}{\text{Bearbeitungszeit der Aufgabe}}$$

Diese Kennzahl wird mit jeder richtigen Rechnung erhöht, wobei schnell bearbeitete Aufgaben mehr Gewicht erhalten. Falsche Rechnungen erhöhen den Wert nicht

- **KLT2:** Länge in Sekunden der Einzelaufgabe, die für die Berechnung am längsten dauerte
- **KLT3:** Länge in Sekunden vom Start der KLT-Applikation, bis die erste Aufgabe bearbeitet ist
- **KLT4:** Durchschnittliche Zeit aller bearbeiteten Aufgaben (egal ob richtig oder falsch)
- **KLT5:** Durchschnittliche Zeit aller richtig beantworteten Aufgaben
- **KLT6:**

$$\frac{\text{Zeit aller Richtigen}}{\text{Durchschnittliche Zeit aller Falschen}}$$
- **KLT7:**

$$\frac{\text{Anzahl der richtig Beantworteten}}{\text{Anzahl der falsch Beantworteten}}$$
- **KLT8:** Gesamtzeit, bis erste Aufgabe richtig beantwortet ist
- **KLT9:** Anzahl Fehler
- **KLT10:** Zeit in Sekunden von der Signalisation eines Events, bis die erste Aufgabe angezeigt wird

Die KLT-Kennzahlen werden jeweils für jeden Event individuell berechnet. Nach der Berechnung der KLT-Werte wurden diese mit den zum gleichen Event dazugehörigen Daten aus dem Matlab zusammengebracht, um die Analyse durchführen zu können. In der Abbildung 4.1 kann man eine schematische Darstellung des Datenaufbereitungs- und Analyseprozesses sehen.

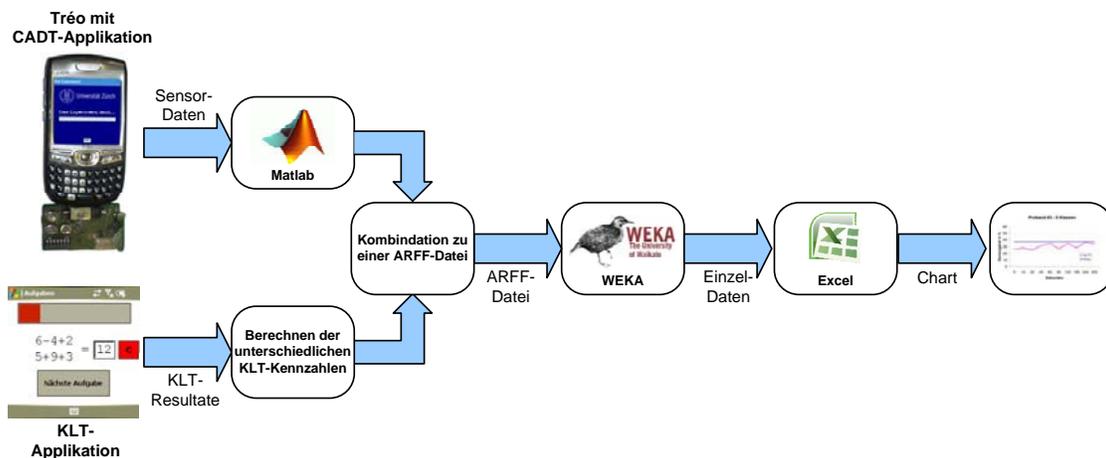


Abbildung 4.1: Schematischer Ablauf der Datenaufbereitung und Analyse

4.2 Deskriptive Statistik

Um eine Vorhersage über den optimalen Unterbrechungszeitpunkt machen zu können, mussten die beim Experiment gesammelten Daten klassifiziert werden. Wir verwendeten dafür das WEKA Data Mining Software Package² Version 3.4. Für die Klassifizierung der Daten wurde der Lernalgorithmus "J48" eingesetzt, mit dessen Hilfe wir Entscheidungsbaume generieren konnten (siehe Listing 4.1). Dieser WEKA-Algorithmus basiert auf dem von J. Quinlan entwickelten C4.5 Maschine Learning Algorithmus [29]. Um mit WEKA beispielsweise eine Vorhersage auf die 5 Stufen des Unterbrechungslevels machen zu können, mussten vor der Berechnung des Entscheidungsbaums, die nicht benötigten Kennzahlen (KLT1 bis KLT10) aus dem Datenbestand gefiltert werden. Danach mussten die 5 Stufen im WEKA diskretisiert werden. Bei der Berechnung des J48-Decision Trees wurde zusätzlich noch das 10-Kreuzvalidierungsverfahren angewendet, um dem Overfitting entgegen zu wirken. Mit gleichem Vorgehen, wurden alle weiteren Vorhersagen aus den Sensordaten auf die entsprechenden Kennzahlen (Unterbrechungslevel, KLTs) erstellt. Das Resultat daraus ist nebst dem Entscheidungsbaum eine Confusion- oder auch Wahrheitsmatrix. In dieser Matrix kann nun abgelesen werden, welche Klassifizierung der Daten, mit Hilfe des Entscheidungsbaums, gemacht worden ist und ob sie richtig beziehungsweise falsch gewesen ist. Die Ergebnisse werden dann in *richtig/falsch-positiv/negative* eingeteilt. Gleichzeitig wird dabei auch ein Wert in Prozent berechnet, welcher die Genauigkeit angibt, mit welcher der Entscheidungsbaum die Realität abgebildet hat und wie exakt die Resultate der Leistungstests, anhand der Sensordaten, berechnet werden können.

```

hour_of_day <= 8
| sensor_mean_1 <= 2292.18: '(0.72-0.97)' (2.0)
| sensor_mean_1 > 2292.18: '(0.24-0.48]' (2.0/1.0)
hour_of_day > 8
| sensor_mean_9 <= 1048.249
| | audio_cepstral_1_mean <= -2.7503
| | | gps_speed_mean <= 0.5493: '(0.48-0.72]' (4.0)
| | | gps_speed_mean > 0.5493
| | | | sensor_mean_2 <= 3810.4957
| | | | | sensor_mean_12 <= 1673.775
| | | | | | sensor_std_12 <= 1229.0454
| | | | | | | audio_cepstral_4_std <= 0.040741: '(0.48-0.72]' (7.0)
| | | | | | | audio_cepstral_4_std > 0.040741
| | | | | | | | audio_zero_crossing_rate_mean <= 0.53719
| | | | | | | | | sensor_std_4 <= 9.8518: '(0.48-0.72]' (4.0)
| | | | | | | | | sensor_std_4 > 9.8518: '(0.72-0.97]' (3.0/1.0)
| | | | | | | | | audio_zero_crossing_rate_mean > 0.53719: '(0.72-0.97]' (7.0)
| | | | | | | sensor_std_12 > 1229.0454
| | | | | | | | sensor_std_12 <= 1603.7266
| | | | | | | | | sensor_mean_2 <= 3779.274
| | | | | | | | | | audio_cepstral_2_std <= 0.04782: '(-inf-0.24]' (2.0)
| | | | | | | | | | audio_cepstral_2_std > 0.04782: '(0.24-0.48]' (6.0)
| | | | | | | | | | sensor_mean_2 > 3779.274: '(0.72-0.97]' (2.0)
| | | | | | | | sensor_std_12 > 1603.7266

```

Listing 4.1: Exemplarischer Ausschnitt eines mit dem J48-Algorithmus erzeugten Entscheidungsbaums

²<http://www.cs.waikato.ac.nz/ml/weka/index.html>

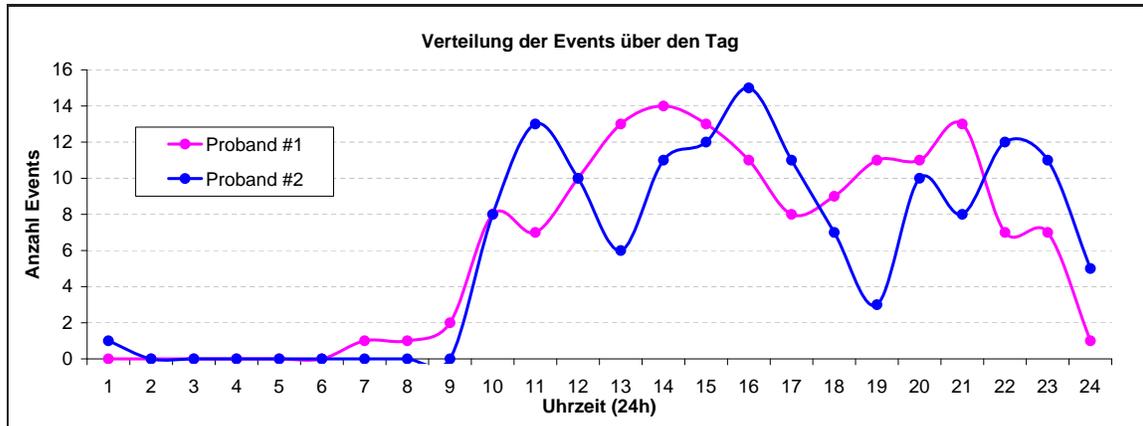


Abbildung 4.2: Verteilung der Events der Probanden #1 und #2

Bei unserem Experiment haben, wie bereits erwähnt, 5 Testpersonen teilgenommen. Insgesamt wurde dabei ungefähr 15 GB Datenmaterial gesammelt. Der Test dauerte im Durchschnitt 25,4 Tage und es wurden 894 Events auf den Geräten aufgezeichnet, wobei 750 (= 84%) beantwortet wurden. In der Abbildung 4.2 sieht man zwei Graphen, welche die zeitliche Verteilung der beantworteten Events, auf einen Tag verteilt, aufzeigen. Es ist gut zu erkennen, dass die Probanden die Events mehrheitlich von morgens 8:00 Uhr bis abends 23:00 Uhr bearbeitet hatten. Es ist interessant, dass es bei Proband #2 zwei markante Einbrüche um 13:00 und 19:00 Uhr gibt. Die Erklärung dafür ist, dass die Testperson das Gerät jeweils während des Mittagessens nicht dabei hatte und so keine Events beantwortete. Die zweite Reduktion der Events war jeweils zum Zeitpunkt der Heimreise des Probanden, während er auch keine Events beantworten konnte. Im Gespräch hat die Testperson gesagt, dass sie während diesen zwei Situationen auch keine Mobiltelefonate entgegen nimmt, womit das "natürliche" Verhalten der Testperson gegeben ist, und dies keine Ausnahme darstellt.

Der leichte Einbruch der Anzahl Events beim Probanden #1 (um 17:00 Uhr) lässt sich durch die Laufzeit des Akkus erklären. Dadurch, dass die Gassensoren einen sehr hohen Stromverbrauch haben, wurden sie jeweils erst 15 Minuten vor einem Event eingeschaltet. Wenn nun die Akkuleistung bereits stark nachgelassen hat (nach etwa 4-6 Stunden) und 15 Minuten vor einem anstehenden Event die stromhungrige Gassensorik eingeschaltet wurde, kam es wiederholt vor, dass kurz vor dem eigentlichen Event die ganze Energie aufgebraucht war und das Gerät ausschaltete. Nach dem Austauschen des Akkus begann die Zeitberechnung für einen neuen Event von vorne, um auszuschließen, dass die Testperson den Event "erwartet".

Bei der Betrachtung der unterschiedlichen Unterbrechungslevel, welche über den ganzen Tag verteilt sind, kann man sehr gut erkennen, zu welchen Zeiten sich die Testperson eher durch einen Event gestört gefühlt hat, und wann die Zeitpunkte eher angenehmer gewesen sind. Die Abbildung 4.3 veranschaulicht, dass die roten (Level 5) und orangen (Level 4) Störlevel am Mor-

gen bis zum Mittag eher dominant sind und dann gegen Abend abnehmen. Gleichzeitig sind die grünlichen Bereiche (Level 1 & 2) während eines Arbeitstags selten aufgetreten, wobei sie gegen Abend ziemlich grosse Bereiche einnehmen. Ein Vergleich (siehe Abbildung 4.4) mit einem anderen Probanden zeigt, dass die Verteilung der hohen Störlevel sehr ähnlich sind und die grünlichen Bereiche jeweils am Nachmittag/Abend stärker ausgeprägt auftreten.

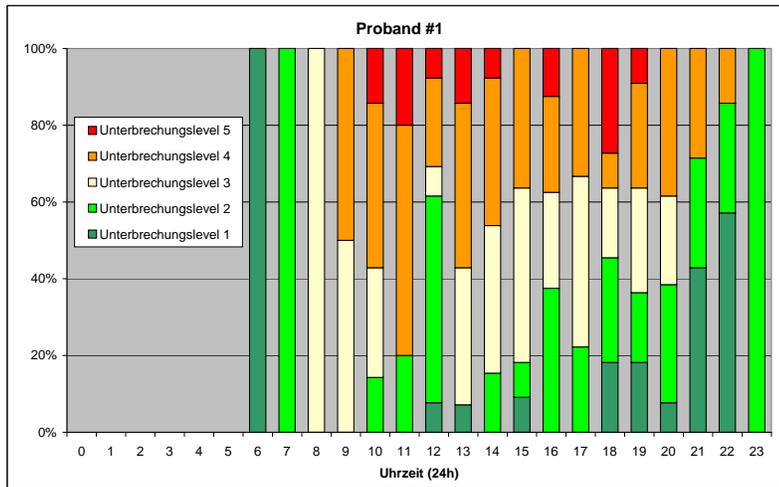


Abbildung 4.3: Vorkommen der unterschiedlichen Unterbrechungslevel über den Tag verteilt - Proband #1

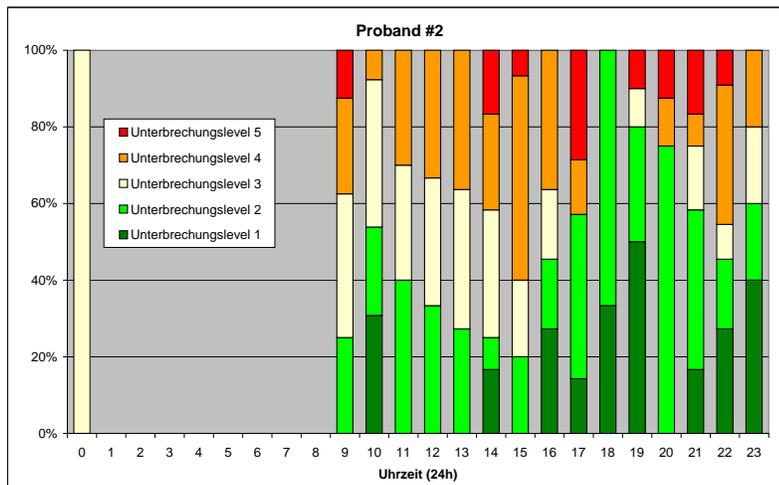


Abbildung 4.4: Vorkommen der unterschiedlichen Unterbrechungslevel über den Tag verteilt - Proband #2

4.2.1 Unterbrechungsstufe

Im Gegensatz zu den vielen Arbeiten, die im Vorfeld von diesem Experiment zu ähnlichen Themen geschrieben worden sind (siehe Kapitel 2), war bei dieser Arbeit der von den Testpersonen selbst empfundene Unterbrechungslevel nicht von primärem Interesse. Wir haben die Werte aber trotzdem bei jedem Event erhoben, um herauszufinden, ob es einen signifikanten Zusammenhang zwischen dem empfundenen Unterbrechungslevel und der Leistungsfähigkeit gibt.

Allgemeine aussagen

In der Abbildung 4.5 kann man die Anzahl der gewählten Unterbrechungslevel ablesen. Von den knapp 750 Einzelwerte waren zweimal 27% im 2. und 3., 22% im 4., 14% im 1. und 9% im 5. Bereich eingeordnet worden (siehe Tabelle 4.1). Der Durchschnittswert aller unterschiedlichen Werte liegt bei 2.83 was ziemlich genau dem mittleren Unterbrechungslevel entspricht. Es kann also generell gesagt werden, dass sich die Testpersonen im Durchschnitt durch eine Unterbrechung weder "stark gestört" noch "nicht gestört" gefühlt haben.

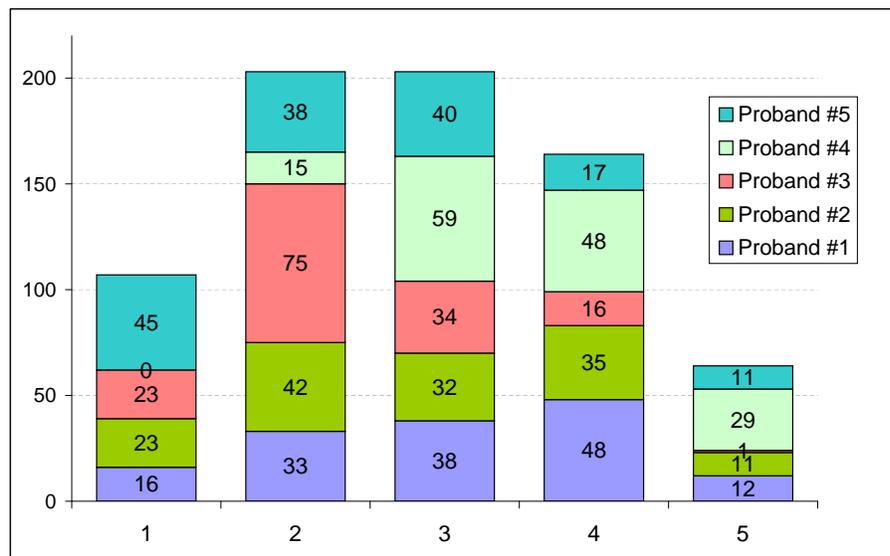


Abbildung 4.5: Übersicht über der Eventverteilung aller Probanden auf die Unterbrechungslevel 1 bis 5

Die Berechnung der Korrelationskoeffizienten zwischen dem Unterbrechungslevel und den unterschiedlichen KLT-Kennzahlen³ hat gezeigt, dass es nur einen minimalen Zusammenhang zwischen diesen Werten gibt. Obwohl erwartet wurde, dass die unterschiedlichen Varianten des KLT einen Zusammenhang mit dem Unterbrechungslevel haben, konnte gezeigt werden, dass beispielsweise beim Probanden #1 die Korrelation nach Pearson beim KLT1 = -0.27, KLT4 = 0.22

³Es hat sich herausgestellt, dass die KLT1, KLT4 und KLT5 die beste Aussagekraft für die Messung der Leistungsfähigkeit besitzen.

Unterbrechungslevel	Anzahl	%
1	107	14.4
2	203	27.3
3	203	27.3
4	164	22.1
5	64	8.6
Summe	741	100

Tabelle 4.1: Verteilung aller Events auf die einzelnen Unterbrechungslevel

und $KLT5 = 0.13$ beträgt, wobei dies jeweils auf dem Niveau von 0.01 (2-seitig) signifikant ist. Im weiteren Verlauf dieser Arbeit, werden nur noch die Kennzahlen $KLT1$, $KLT4$ und teilweise der $KLT5$ als Index für die Leistungsfähigkeit verwendet. Es hat sich herausgestellt, dass diese Werte die Leistung der Testpersonen zum Zeitpunkt eines Events am besten abbilden. Diese Kennzahlen berücksichtigen jeweils die Anzahl der richtigen Aufgaben sowie die Bearbeitungszeit. Im $KLT4$ werden im Vergleich zum $KLT5$ auch die falsch gerechneten Aufgaben miteinbezogen, da nach H. Düker *„Auch eine falsche Lösung ein gewisses Mass an Konzentration erfordert und somit positiv zu werten ist. Wenn man von der Möglichkeit des blossen Ratens absieht“* [9].

In den Abbildungen 4.6 kann man erkennen, dass man bei genauer Betrachtung einen leichten Trend der Datenpunkte von links oben, nach rechts unten gibt. Die eingezeichnete Regressionsgerade bestätigt zusätzlich diese Vermutung. Die beste Korrelation, des Probanden #1 (nur -0.27) sagt allerdings aus, dass es nur einen minimal negativen Zusammenhang zwischen dem Störlevel und der Kennzahl $KLT1$ gibt. In den Abbildungen 4.7 und 4.8 ist allerdings kein wirklich eindeutiger Trend ersichtlich.

Interessant ist, dass auch bei Probanden #2, #4, #5 nur sehr schwache Korrelationen berechnet werden konnten. Bei Proband #3 gibt es allerdings eine bessere Korrelation zwischen dem Unterbrechungslevel und den KLT-Werten. Die Korrelationen betragen für $KLT1 = -0.40$, $KLT4 = 0.37$ und $KLT5 = 0.35$. Diese deutlich besseren Korrelationen zwischen den Kennzahlen ist auch auf den Darstellungen 4.9, 4.10 und 4.11 anhand der Regressionsgerade erkennbar. Sie weisen eine deutlich markantere Neigung, im Vergleich zu jener von Proband #1 auf.

Damit die Resultate eindeutig gewesen wären, hätten die KLT-Werte in den Randbereichen des Unterbrechungslevels (1 und 5) deutlich höher oder niedriger sein sollen. Wie die Vorhersagen auf diese Werte anhand der Sensordaten ist, wird im Abschnitt 4.2.2 behandelt.

Es stellt sich die Frage, weshalb die KLT-Werte bei 4 von 5 Probanden keine eindeutige, und bei einer einzelnen Person eine nur mittel-starke Korrelation aufweisen. Eine mögliche Erklärung wäre, dass die KLT-Kennzahlen die Leistungsfähigkeit der Testpersonen zu wenig genau erfasst und deshalb mit steigendem Störlevel nicht sinken ($KLT1$), oder ansteigen ($KLT4$ & $KLT5$). Eine andere mögliche Erklärung könnte die Tatsache sein, dass die Leistungsfähigkeit nicht vom empfundenen Störlevel abhängt. Eine genauere Betrachtung dieser Fragestellung kann im Abschnitt 5.1 nachgelesen werden.

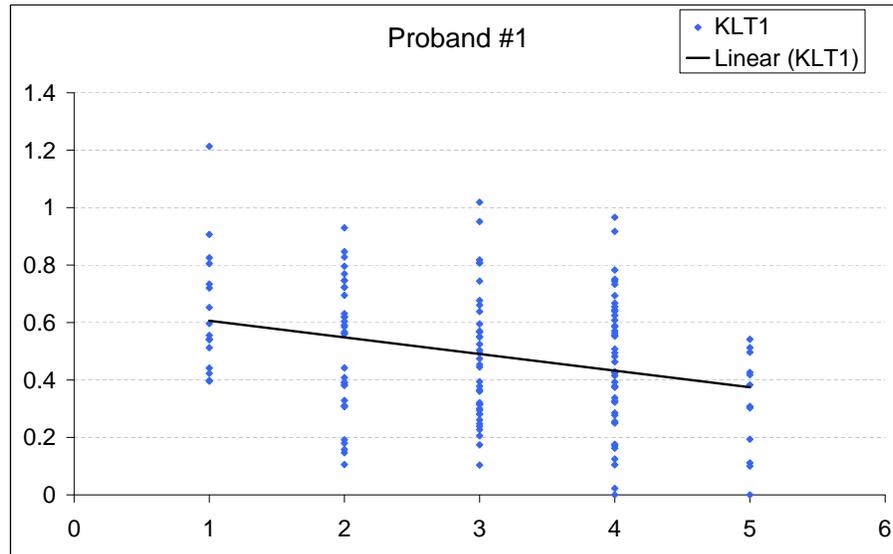


Abbildung 4.6: Korrelation mit Regressionsgerade zwischen Unterbrechungslevel und KLT1-Resultate - Proband #1 - Hohe KLT1-Werte $\hat{=}$ gute Leistungsfähigkeit

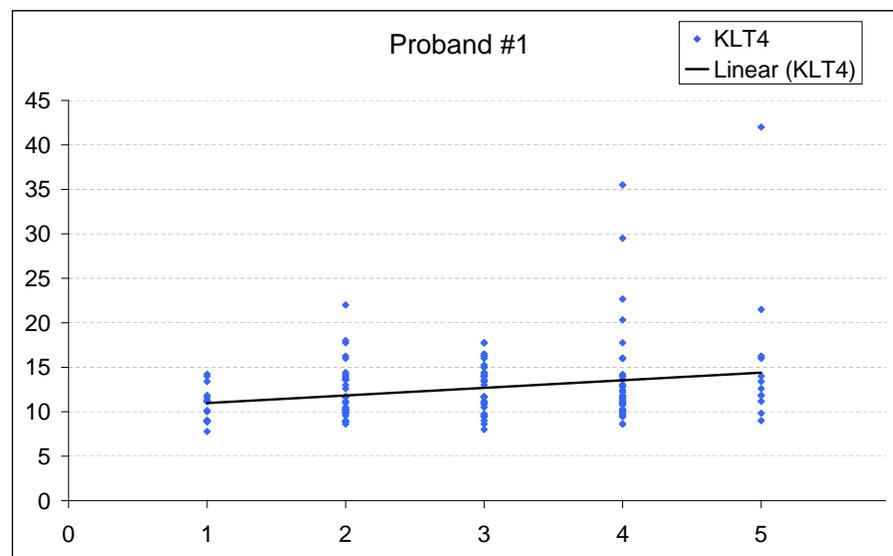


Abbildung 4.7: Korrelation mit Regressionsgerade zwischen Unterbrechungslevel und KLT4-Resultate - Proband #1 - Tiefe KLT4-Werte $\hat{=}$ gute Leistungsfähigkeit

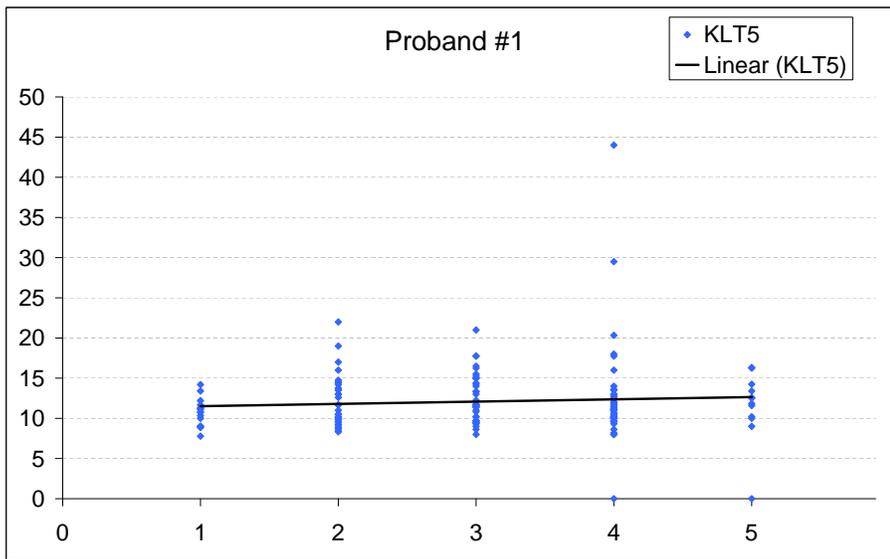


Abbildung 4.8: Korrelation mit Regressionsgerade zwischen Unterbrechungslevel und KLT5-Resultate - Proband #1 - Tiefe KLT4-Werte $\hat{=}$ gute Leistungsfähigkeit

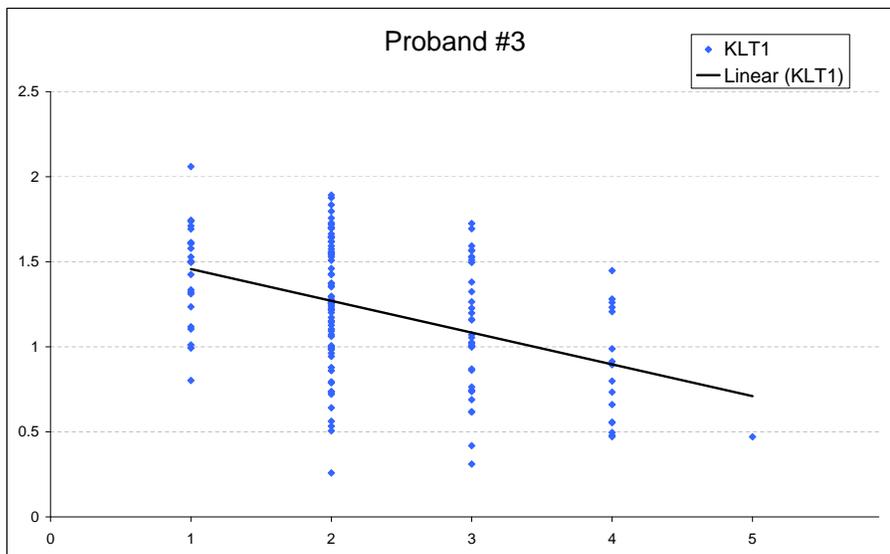


Abbildung 4.9: Korrelation mit Regressionsgerade zwischen Unterbrechungslevel und KLT1-Resultate - Proband #3

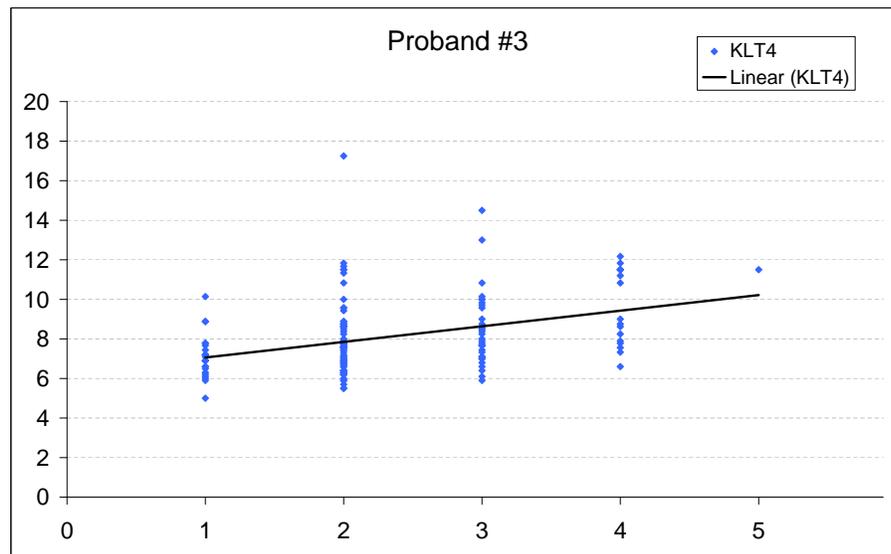


Abbildung 4.10: Korrelation mit Regressionsgerade zwischen Unterbrechungslevel und KLT4-Resultate - Proband #3

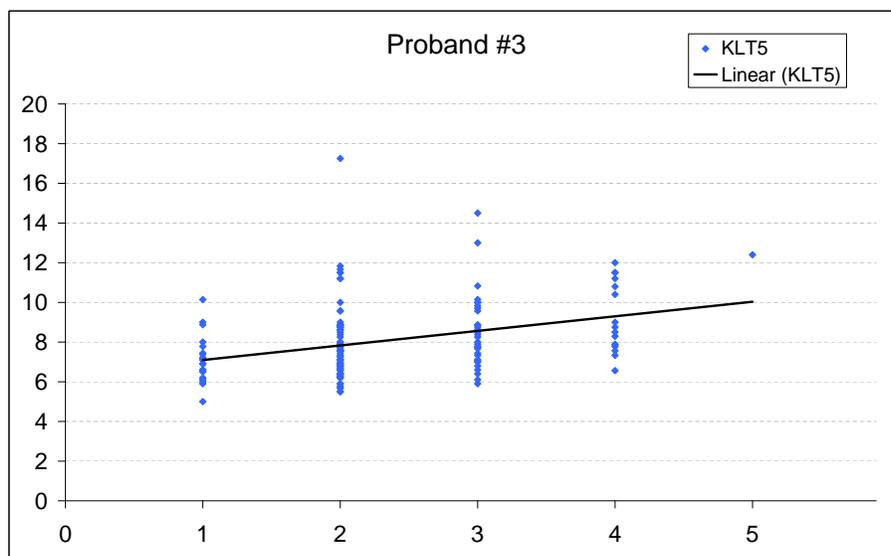


Abbildung 4.11: Korrelation mit Regressionsgerade zwischen Unterbrechungslevel und KLT5-Resultate - Proband #3

4.2.2 Vorhersage der Leistungsfähigkeit mit Sensordaten

In diesem Abschnitt werden die eigentlichen Resultate behandelt, welche im Experiment herausgefunden werden sollten. Es wird gezeigt, mit welcher Genauigkeit es möglich ist, anhand der Sensordaten, die mit dem Tréo aufgezeichnet wurden, Rückschlüsse auf die aktuelle Leistungsfähigkeit einer Person zu ziehen. Da der Vergleich zwischen den Unterbrechungslevel und den KLT-Werte keinen eindeutigen Zusammenhang hervorbrachte, musste nach einer Möglichkeit gesucht werden, wie die KLT-Resultate für das Machine Learning klassifiziert werden konnten. Der J48-Algorithmus benötigt zwingend eine Einteilung in Klassen, um Regeln erstellen zu können, anhand derer aus den Sensordaten auf die Leistungsfähigkeitsklasse geschlossen werden kann. Wir haben uns entschieden zwei unterschiedliche Klassifizierungen vorzunehmen. Bei der einen werden die KLT-Werte in zwei und bei der anderen in fünf Kategorien eingeteilt. Für den Vergleich, wurden auch die Vorhersagen auf die Unterbrechungslevel gemacht, welche bereits in 5 Klassen eingeteilt sind.

2-Klassen Vorhersage

Da es bei den KLT-Kennzahlen keine scharfe Grenze gibt, ab welcher eine Leistung schlecht oder gut ist, wurden die KLT-Werte folgendermassen in zwei Bereiche eingeteilt. Im ersten Teil sind alle mittleren bis sehr guten Resultate enthalten. Im zweiten Teil sind die eher schlechten Leistungen eingeordnet. Da im Durchschnitt etwa 1/5 alle Events als störend (Unterbrechungslevel 5 & teilweise 4)⁴ klassifiziert worden sind, haben wir auch etwa 30 KLT-Werte als *schlecht* eingestuft. Bei der Suche dieser Abgrenzungen haben die Graphen 4.12a, 4.12b, 4.13a, 4.13b, 4.14a, 4.14b, der unterschiedlichen KLT-Resultate zusätzliche Hinweise gegeben. Auf ihnen ist gut sichtbar, an welcher Stelle die Werte sinken und demnach die Leistungsfähigkeit nachlässt. Die blaue Funktion bildet dabei die Summe aller Events verteilt auf die dazugehörigen KLT-Werte. Die deutlich tieferen farbigen Funktionen sollen die Zugehörigkeit zu den Unterbrechungslevel aufzeigen. Wäre der Zusammenhang zwischen den KLT's und dem Unterbrechungslevel grösser, würde jede diese tieferen Kurven an der entsprechenden Stelle ein Maximum aufweisen.

Nach der Bestimmung der Grenzen zwischen den zwei Bereichen, konnten die Sensordaten mit Hilfe des WEKA verarbeitet werden und die Vorhersagen für die KLT1 und KLT4 wurden durch den J48-Algorithmus berechnet. Die Resultate von S. Christiani [4] haben gezeigt, dass die besten Vorhersagen entstehen, wenn für die Analyse, Sensordaten von 70 Sekunden benutzt werden. In unserer Auswertung werden allerdings trotzdem alle 10 Zeitintervalle ausgerechnet, wobei die Detailanalysen nur auf dem 90 Sekunden Datenmaterial basieren.

⁴Es wurden bei allen Versuchspersonen die Anzahl der Events im schlechtesten Unterbrechungslevel und die halbe Anzahl im zweitschlechtesten Unterbrechungslevel zusammengezählt. Dies ergab ca. 30

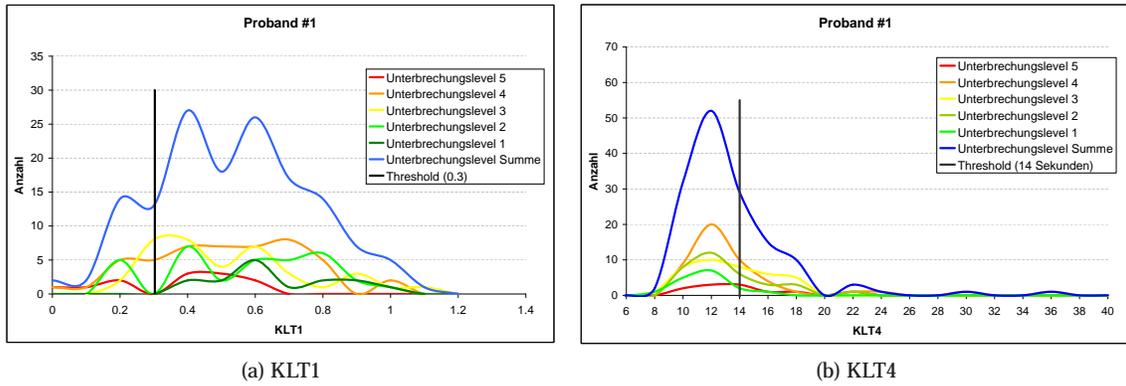


Abbildung 4.12: Häufigkeiten der KLT1- & KLT4-Resultate für Proband #1 mit Einordnung in die Unterbrechungslevel

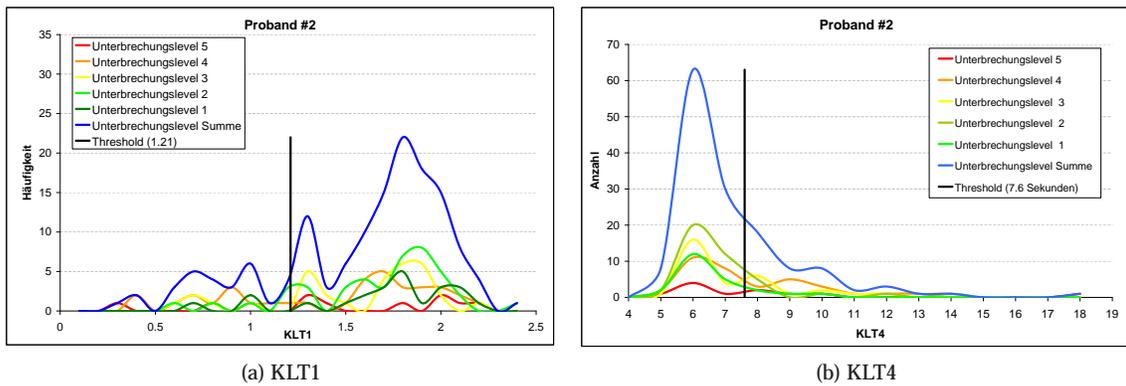


Abbildung 4.13: Häufigkeiten der KLT1- & KLT4-Resultate für Proband #2 mit Einordnung in die Unterbrechungslevel

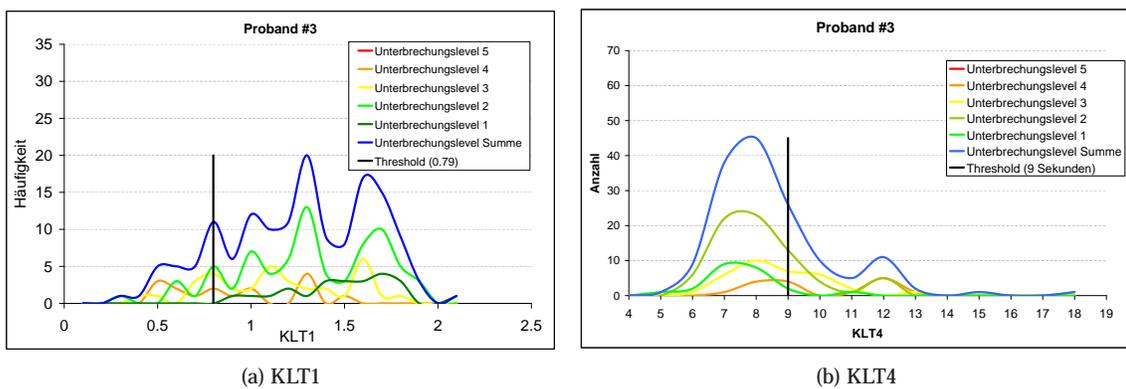


Abbildung 4.14: Häufigkeiten der KLT1- & KLT4-Resultate für Proband #3 mit Einordnung in die Unterbrechungslevel

Zu unserem Bedauern sind die berechneten Resultate sehr unbefriedigend. Obwohl die Vorhersagegenauigkeit über 60% liegt, übersteigt sie nur gerade beim Probanden #3 zu einem Zeitpunkt den Prior⁵. Die Abbildungen 4.15, 4.16 und 4.17 belegen dieses Erkenntnis noch zusätzlich. Obwohl die Werte nur ganz knapp unterhalb des Priors liegen, würde eine zufällige Klassifizierung in *leistungsfähige* oder *weniger leistungsfähige* Zeitpunkte, statistisch gesehen, ein besseres Resultat bedeuten. Die Wahrheitsmatrix Tabelle 4.2 von Proband #1 für die KLT4 bei 90 Sekunden zeigt auch ein eindeutiges Bild. Von den insgesamt 148 verarbeiteten Events (115 im *guten* KLT4 Bereich, 33 im *schlechten* Bereich) werden durch den J48-Algorithmus nur 98 richtig-positiv und 4 richtig-negativ klassifiziert. 29 Events werden als *gut* eingeordnet, obwohl sie eigentlich *schlecht* wären. Mit 17 passierte das entsprechend Umgekehrte. Für den KLT1 wurden 9 als richtig-positiv und 87 richtig-negativ eingestuft (siehe Tabelle 4.3).

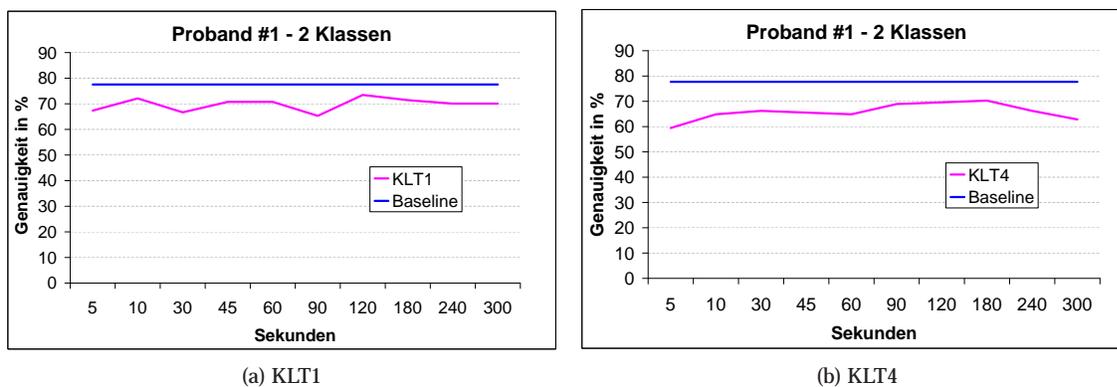


Abbildung 4.15: Vorhersagen für Proband #1

KLASSIFIZIERT ALS ->	a	b	WIRKLICHKEIT
	98	17	a = 0.0 - 14.0
	29	4	b = 14.0 - 34.8

Tabelle 4.2: Vorhersagen des KLT4 (2-Klassen mit Threshold 14.0) für Proband #1 - 90 Sekunden Sensordaten - Genauigkeit 68.9%

KLASSIFIZIERT ALS ->	a	b	WIRKLICHKEIT
	9	24	a = 0.0 - 0.3
	27	87	b = 0.3 - inf

Tabelle 4.3: Vorhersagen des KLT1 (2-Klassen) für Proband #1 - 90 Sekunden Sensordaten - Genauigkeit 65.3%

⁵Der Prior ist der Grenzwert, bis zu welchem der Zufall eine bessere Vorhersage macht als die Werte, welche darunter liegen. Der Prior wird berechnet, indem die Anzahl der Werte, in der am häufigsten vorkommenden Klasse, durch die Summe der Werte über alle Klassen dividiert wird. Bsp.: Zwei Klassen mit 30 und 60 Werten (Summe daraus = 90). Der Prior ist dann $60 : 90 = 66.6\%$

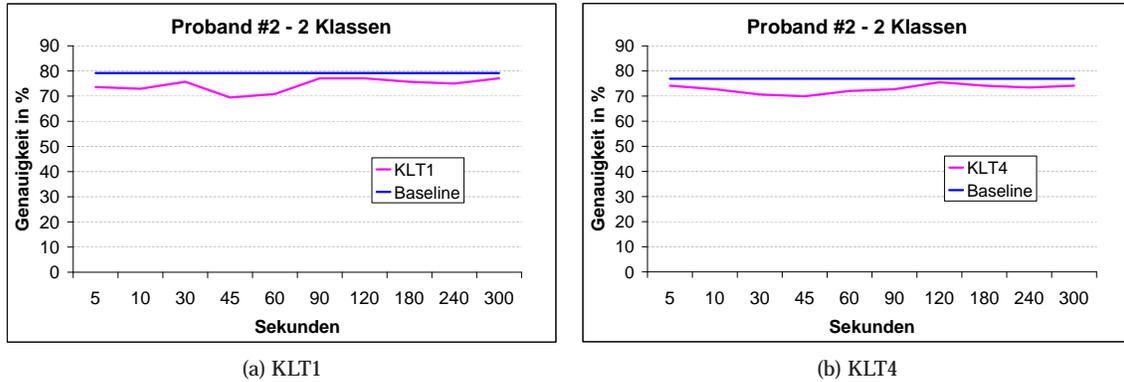


Abbildung 4.16: Vorhersagen für Proband #2

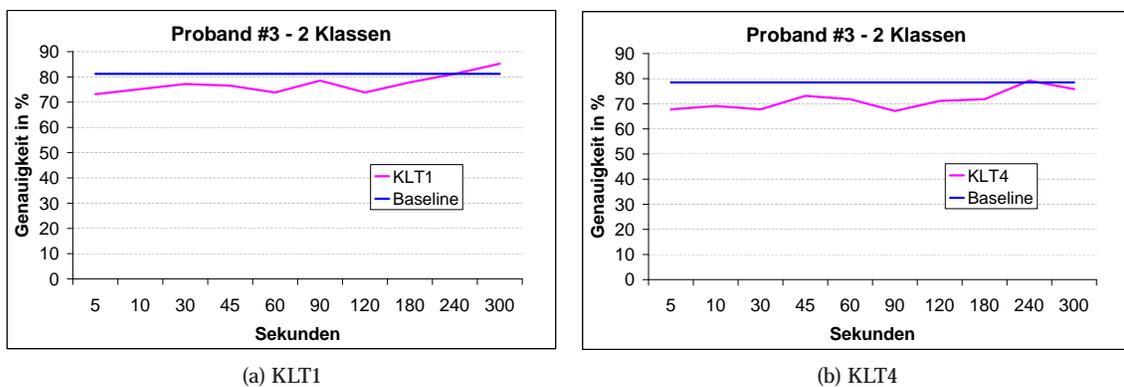


Abbildung 4.17: Vorhersagen für Proband #3

Der Schwellenwert zwischen den *guten* und *schlechten* Events ist beim KLT4 = 14.0 und beim KLT1 = 0.303,

5-Klassen Vorhersage

Als Variante zur 2-Klasseneinteilung haben wir Berechnungen mit einer 5er Einteilung durchgeführt. Die 5 Stufen wurden gewählt, da die Vorhersage des Unterbrechungslevels dieselbe Anzahl besitzt und für einen Vergleich herangezogen werden kann. Bei der Betrachtung der Vorhersagen (siehe Abbildungen 4.18, 4.19, 4.20) für die 5 Unterbrechungslevel sieht man, dass auch diese Resultate nur an wenigen Stellen den Prior überragen. Obwohl bei Proband #2 die Vorhersage, mit 120 Sekunden Datenumfang, 4% über dem Zufall liegt, genügt ein solcher minimaler Ausschlag nicht als verlässliche Vorhersage. Die Tabelle 4.4 stellt die dazugehörige Wahrheitsmatrix dar.

Zusätzlich zu dem Unterbrechungslevel wurden auch die Vorhersagen der KLT1 in 5 Klassen eingeteilt. Dabei verwendeten wir bei der Testperson #1 die gleichmässig verteilten Intervalle des

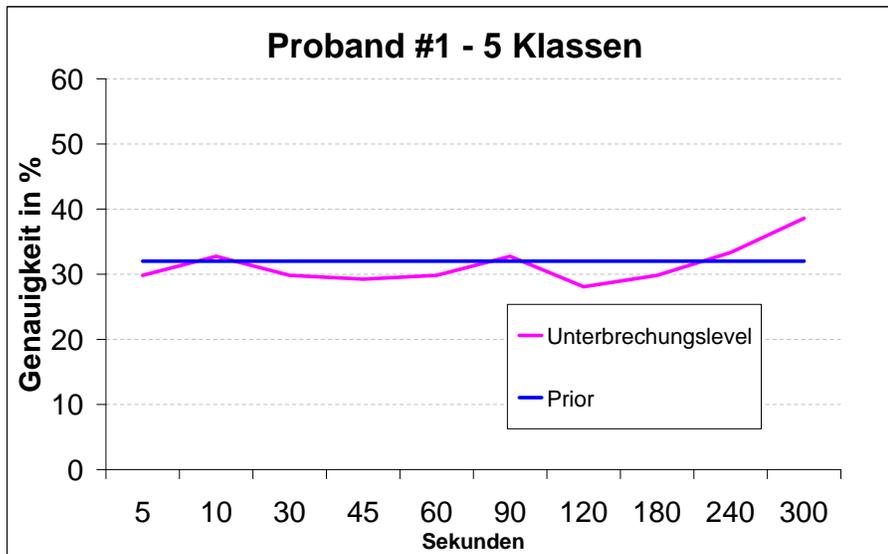


Abbildung 4.18: Vorhersagen der Unterbrechungslevel (5-Klassen) für Proband #1

KLASSIFIZIERT ALS ->	a	b	c	d	e	WIRKLICHKEIT
	16	13	4	7	1	a = 1
	7	19	7	8	2	b = 2
	7	7	8	9	1	c = 3
	3	8	12	9	3	d = 4
	3	4	1	0	3	e = 5

Tabelle 4.4: Vorhersagen der Unterbrechungslevel (5-Klassen) für Proband #2 - 120 Sekunden Sensordaten - Genauigkeit 33.9%

KLT1 von 0.24 / 0.48 / 0.72 / 0.97. Die daraus entstandene Wahrheitsmatrix und Grafik sieht man unter Tabelle 4.5 und Abbildung 4.21. Die Abbildungen 4.22 und 4.23 wurden zusätzlich für Proband #2 und #3 erstellt, wobei auch eine symmetrische Einteilung der Klassen erfolgt ist. Auch diese Vorhersagen reichen nur ganz vereinzelt über den Prior hinaus und sind demnach für eine verlässliche Vorhersage genauso ungeeignet.

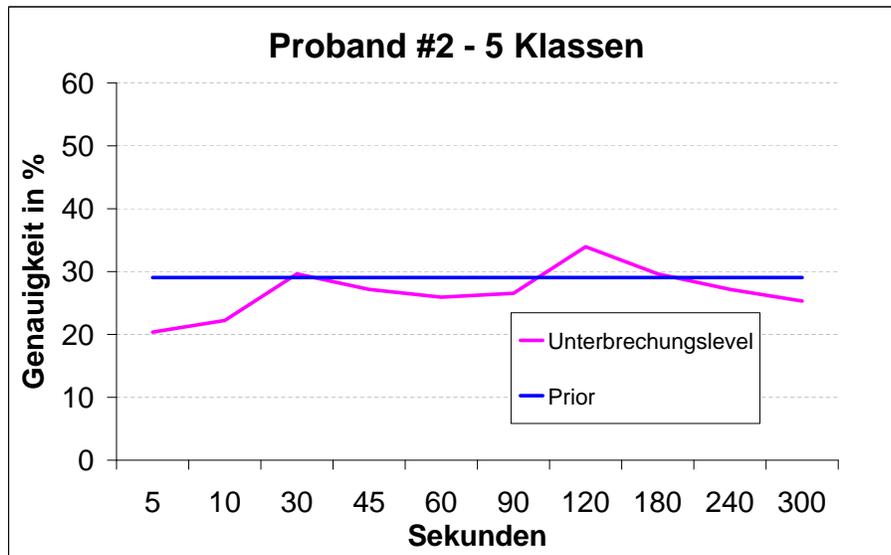


Abbildung 4.19: Vorhersagen der Unterbrechungslevel (5-Klassen) für Proband #2

KLASSIFIZIERT ALS ->	a	b	c	d	e	WIRKLICHKEIT
4	8	7	2	0		a = 0.00 - 0.24
5	25	16	6	0		b = 0.24 - 0.48
8	12	22	6	1		c = 0.48 - 0.72
2	9	8	4	0		d = 0.72 - 0.97
1	1	0	0	0		e = 0.97 - inf

Tabelle 4.5: Vorhersagen des KLT1 (5-Klassen) für Proband #1 - 180 Sekunden Sensordaten - Genauigkeit 37.4%

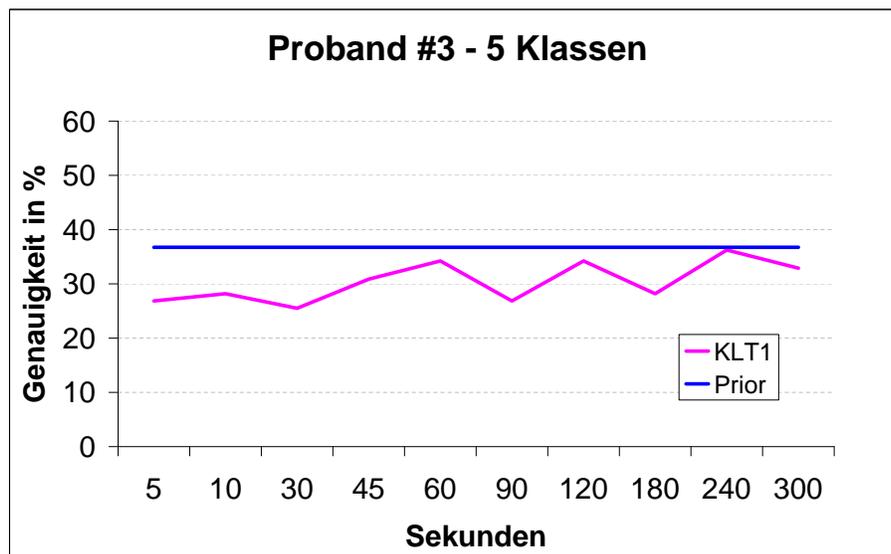


Abbildung 4.20: Vorhersagen der Unterbrechungslevel (5-Klassen) für Proband #3

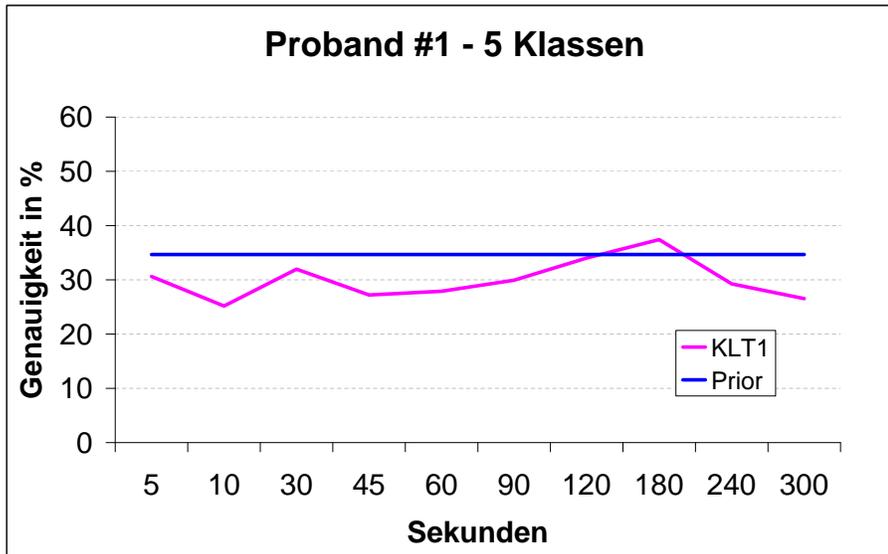


Abbildung 4.21: Vorhersagen der KLT1-Werte (5-Klassen) für Proband #1

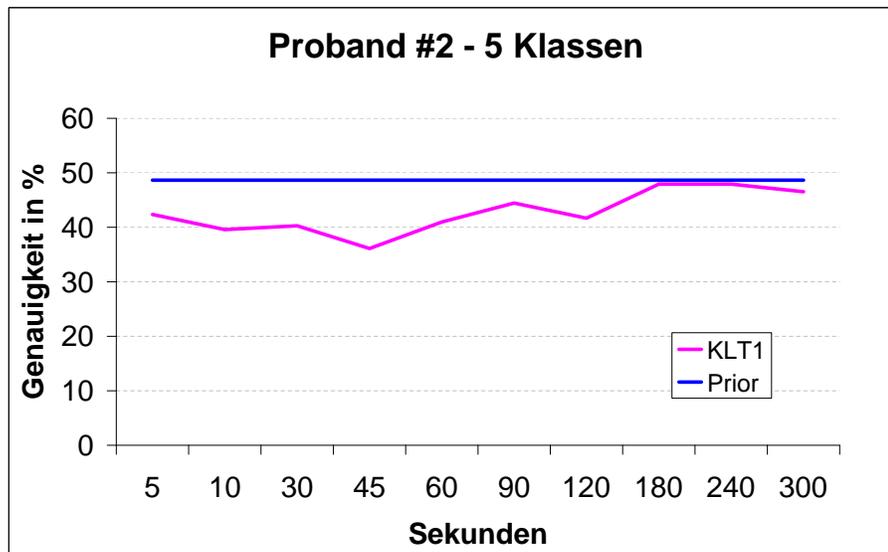


Abbildung 4.22: Vorhersagen der KLT1-Werte (5-Klassen) für Proband #2

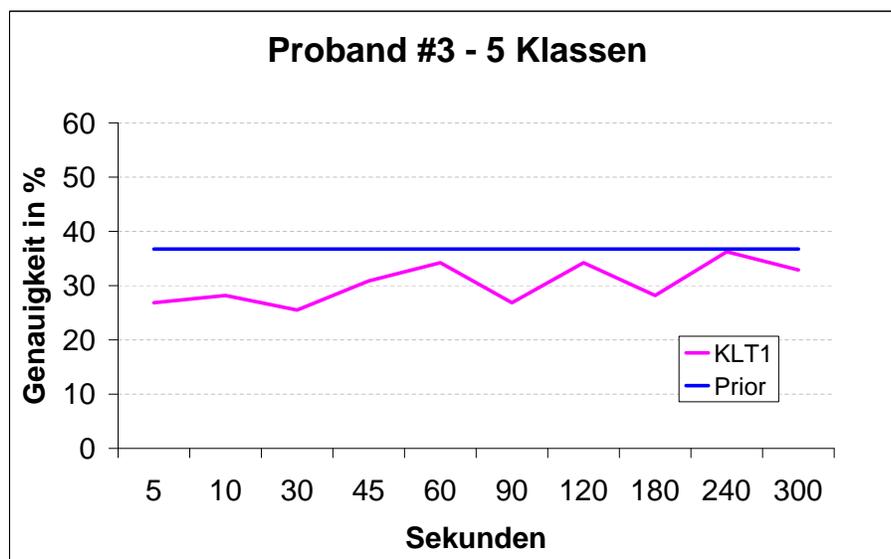


Abbildung 4.23: Vorhersagen der KLT1-Werte (5-Klassen) für Proband #3

4.2.3 Interview mit den Probanden

Am Ende des Experiments wurde mit allen Probanden ein Interview durchgeführt. Es sollten dabei einige Fragen beantwortet werden, die bei der Analyse der Resultate aufgetaucht sind. Das Komplette Interview kann im Anhang A nachgelesen werden.

Erkenntnisse aus den Interviews

Die Auswertung der Gespräche hat einige interessante Erkenntnisse hervorgebracht. Beispielsweise haben Probanden, welche bei den Events beinahe immer alle KLT-Aufgaben gelöst haben, angegeben, dass sie sehr schnell Rechnungen lösen können. Diese deckt sich auch mit den Beobachtungen aus den originalen KLT, bei welchem die schnellen "Rechner" mehr bearbeitet haben. Bei der Frage, ob sie einen gewissen Leistungsdruck beim Lösen der Aufgaben empfunden haben, hat die Mehrheit mit ja geantwortet. Dies könnte auch eine Erklärung dafür sein, weshalb die KLT-Werte relativ gleichbleibend stabil waren. Es wurde auch immer versucht, alles anstehende Aufgaben zu lösen.

Eine eher negative Tatsache ist, dass sich die Testpersonen oftmals in der gleichen Umgebung aufgehalten haben. Die Mehrheit eines Arbeitstags wurde am Pult sitzend verbracht und es wurden so nur wenig neue Umgebungssituationen aufgezeichnet. Da dies aber ein normales Verhaltensmuster bei der Arbeit ist, sollte dies nicht kritisiert werden.

5

Diskussion

In diesem Kapitel werden die Resultate aus dem durchgeführten Experiment analysiert und besprochen. Es wird versucht die Zusammenhänge zwischen den Resultaten zu ergründen und gegebenenfalls Ursachen für mögliche Schwachpunkte herzuleiten.

5.1 Zusammenhang Unterbrechungslevel und Konzentrations- Leistungstest

Bei der Analyse der Daten (Abschnitt 4.2.1) wurde herausgefunden, dass der Zusammenhang zwischen dem, von den Testpersonen eingegebenen Unterbrechungslevel und den beim Lösen der KLT-Rechnungen erreichten Resultaten, nur sehr minimal ist. Die Korrelationskoeffizienten die am ehesten einen Zusammenhang anzeigen, betragen nur gerade -0.40 und -0.27 . Die Werte der anderen Probanden betragen -0.20 , -0.08 und 0.02 . Obwohl das Gefühl vermuten lässt, dass es einen Zusammenhang zwischen diesen beiden Werten gibt, haben die Analysen gezeigt, dass dies nicht ohne weiteres bestätigt werden kann.

Bei der Suche nach einer Erklärung für diese Feststellung, haben wir eine mögliche Schwachstelle aufgefunden. Es könnte sein, dass der Leistungstest für unsere Messung nicht geeignet war und wir deshalb keinen Zusammenhang aufzeigen konnten. Um diese Frage zu klären, können die drei Gütekriterien, die ein Test erfüllen muss um als zuverlässig zu gelten, zu Rate gezogen werden. Diese sind *Objektivität*, *Reliabilität* und *Validität* [30]. Diese drei Kriterien waren auch der Grund dafür, dass wir uns für den bekannten KLT entschieden und keinen eigenen Test entwickelt haben. Es stellt sich die Frage, ob durch unsere Anpassungen am ursprünglichen Test ein negativer Einfluss auf die Resultate hervorgerufen wurde. Es müsste deshalb nochmals genau abgeklärt werden, ob die drei Kriterien wirklich erfüllt sind.

Die Objektivität ist in unserem Fall nicht verletzt, da der Test völlig unabhängig vom durchführenden Untersucher ist.

Die Reliabilität, welche sich auf die Zuverlässigkeit der Resultate bezieht, sagt aus, dass der Test

bei Personen mit der gleichen Leistungsfähigkeit auch die gleichen Resultate rauskommen sollten [30]. Wenn die Resultate aus dem Experiment mit den original KLT-Leistungen verglichen werden, die vor dem Beginn des Experiments mit allen Testteilnehmern durchgeführt wurden, kann auch dieses Kriterium als erfüllt angesehen werden. Obwohl bei der Auswertung der Testbogen ein anderes Punktesystem verwendet wurde, sieht man trotzdem, dass sich die Resultate sehr stark gleichen. Die Probanden #2, #3 und #5, die beinahe bei allen Events die Mehrheit der Aufgaben bearbeitet haben, lösten auch entsprechen viele Aufgaben beim originalen KLT. Proband #1 und #4 haben in beiden Varianten des Tests entsprechend weniger bearbeitet. Im Interview (siehe Abschnitt 4.2.3) mit den einzelnen Personen konnte man noch zusätzlich erfahren, dass die drei schnell rechnenden Probanden bestätigten, dass es ihnen generell sehr leicht fällt, Rechungsaufgaben zu lösen.

Bei der Validität geht es um die Inhaltliche Genauigkeit des Tests. Misst dieser auch wirklich das, für was er eingesetzt worden ist. An dieser Stelle könnte möglicherweise das Problem der niedrigen Korrelationen stecken. Der Test wurde ursprünglich entwickelt um die Konzentrationsleistung von Testpersonen zu erheben. In unserem Experiment wurde allerdings nicht die Konzentration des Probanden abgefragt, sondern wie stark die Person im Moment des Events gestört wird. Wir haben demzufolge zwei Kennzahlen miteinander zu kombinieren versucht, die keinen offensichtlichen Zusammenhang haben. Anhand dieser Erklärung, kann daraus geschlossen werden, dass die Resultate durchaus richtig sein können und es wirklich nur einen minimalen Zusammenhang zwischen dem empfundenen Unterbrechungslevel und der Konzentrationsleistung gibt. Diese Vermutung wird auch noch zusätzlich durch die Erkenntnisse aus den Vorhersagen der Unterbrechungslevel (siehe Abbildungen 4.18, 4.19, 4.20) gestützt. Wären die KLT-Resultate generell von schlechter Qualität, würden die vom KLT unabhängigen Vorhersagen für die Unterbrechungslevel möglicherweise besser sein. Da sie allerdings auch nicht genügend gut sind, liegt die Hauptursache womöglich doch nicht am Leistungstest.

5.2 Vorhersagen der Konzentrationsleistung anhand der Sensordaten

Die Berechnung der Konzentrationsleistungs-Vorhersagen der Testpersonen mit Hilfe eines Entscheidungsbaums hat gezeigt, dass dies nicht ohne weiteres möglich ist. Obwohl wir für jeden einzelnen Probanden die Grenzwerte bei denen der KLT-Wert von *gut* nach *schlecht* wechselt, optimal zu bestimmen versucht haben, liegen die Vorhersagen für eine richtige Einordnung in der Wahrheitsmatrix, deutlich unter den Erwartungen. Wie in Kapitel 4.2.2 beschrieben, liegen beinahe alle Berechnungen unter der minimalen Grenze, dem Prior. Aus diesem Grund, würde man mit einer zufälligen Wahl der Klassifizierung besser fahren, als mit unseren berechneten. Um die Fragestellung dieser Arbeit, ob es möglich ist mit den eingesetzten Sensoren die Arbeitsleistung einer Person zu erhöhen, positiv zu beantworten, wäre es allerdings zwingend von Nöten ge-

wesen, dass eine Vorhersage möglich ist. Anhand dieser Klassifizierungen hätte nämlich in der Produktivphase des Experiments entschieden werden sollen, ob ein Event der Testperson "ausgeliefert" wird oder zurückgehalten werden muss, bis ein besserer Zeitpunkt erreicht ist.

Da sowohl die 2-Klassen- wie auch die 5-Klasseneinteilung keine positiven Resultate ergeben haben, musste herausgefunden werden, wo die möglichen Ursachen verborgen sein könnten.

Das Experiment kann man vereinfacht in zwei Hauptbestandteile zerlegen, in denen jeweils das Problem gesucht werden kann. Zum Einen im eingesetzten Test und zum Anderen in die verwendeten Sensoren.

Wie bereits im vorhergehenden Abschnitt erklärt, ist der eingesetzte Konzentrations- und Leistungstest vom Ansatz her richtig gewählt, da er wirklich die Konzentration der Testperson misst. Es kann somit ausgeschlossen werden, dass es Fehler bei der Testkonstruktion gegeben hat. Bei den Sensoren kann ähnlich argumentiert werden. Da die Sensoren bereits bei früheren Experimenten eingesetzt worden sind und damit gute Ergebnisse erzielt worden sind, kann es nicht an der Funktionsweise der Sensoren gelegen haben. Auch die Aufbereitungsart der Daten hat bereits in anderen Berichten gute Arbeit geleistet und sollte deshalb auch keine negativen Ursachen auf das Endresultat haben.

Unser Experiment basiert auf der Annahme, dass die eingesetzten Sensoren anhand von Signalen und Gegebenheiten (Context-Awareness) aus dem Umfeld, Rückschlüsse auf die Konzentration einer Person machen können. Wenn wir davon ausgehen, dass einer dieser zwei Werte sich nicht so verhält, wie wir dies gerne gehabt hätten, würden die Vorhersagen in Mitleidenschaft gezogen werden, wie dies in unserem Fall passiert ist. Es könnte sein, dass die KLT-Werte zwar die Konzentrationsleistung einer Person wiedergeben, sie aber zu wenig stark durch das Umfeld (welches mit den Sensoren aufgezeichnet wird) beeinflusst werden. Vereinfacht gesagt bedeutet dies, dass die Leistung der Testpersonen zwar in einem bestimmten Rahmen schwankt, die Werte aber zu wenig stark ausschlagen, wenn sich die Umweltbedingungen ändern. In dieser Vermutung könnte einer der Hauptgründe, der schlechten Vorhersagen sein. Psychologische Tests werden meistens so konstruiert, dass sie die generelle geistige Leistungsfähigkeit einer Person messen. Diese Leistungsfähigkeit ist für jeden Menschen unterschiedlich und verändert sich über die Zeit nur minimal. Es wird sozusagen die "Grundausstattung" der Hirnleistung gemessen. Genau dies könnte in unserem Experiment einen negativen Einfluss haben. Der KLT misst möglicherweise nur die generelle Leistungsfähigkeit und wird nur schwach durch die akute Stresssituation beeinflusst. Er repräsentiert somit die allgemeine Leistungsfähigkeit und es ist nicht möglich damit eine verlässliche Vorhersage aus den Sensordaten zu machen.

Eine weitere Möglichkeit, welche die KLT-Resultate auch verfälschen könnte, ist der natürliche Ehrgeiz der Probanden. Im Interview mit den Testpersonen (siehe 4.2.3) wurde bestätigt, dass der Test den Ehrgeiz angeregt hat. Jeder Event wurde als kleiner Wettkampf gesehen und die Personen hatten keine Fokussierungsprobleme, was in Stresssituationen die Resultate hätte negativ beeinflussen sollen.

Weiter könnte es möglich sein, dass sich die Konzentration ähnlich wie ein Muskel verhält. Wenn ein Muskel durch das Aufwärmen vorbereitet ist, kann er eine bessere Leistung erbringen als im

kalten Zustand. Unsere Konzentration verhält sich möglicherweise ähnlich. Wenn eine Testperson auf eine Arbeit stark konzentriert ist, kann sie womöglich die ganze Konzentrationsleistung auf einen anstehenden KLT-Event umfokussieren und leistet so eine in dieser Situation verhältnismässig gute Leistung.

Eine weitere Vermutung, welche im Interview allerdings nur durch eine Person bestätigt wurde, war, dass bei einem anstehenden Event, welcher während eines Gesprächs stattfindet, die Gesprächspartner kurz unterbrechen, damit der Test ruhig bearbeitet werden konnte. Der Proband wäre dann nämlich in einer Situation in der er eigentlich keine Unterbrechung wünschen würde, allerdings wären die KLT-Resultate eher gut, da er ohne Ablenkung rechnen konnte.

Ein anderer möglicher Schwachpunkt ist die Kritik an den eingesetzten Sensoren. Obwohl in vorgängigen Arbeiten gezeigt wurde, dass es mit den Sensoren möglich ist, die Aktivität oder die Lokalität einer Person zu erfassen (siehe [3][4]) muss dies nicht unbedingt bedeuten, dass damit auch mögliche Störquellen der Konzentration eruiert werden können. Es könnte sein, dass diese physikalischen Sensoren den Kontext einer Situation zu wenig genau erfassen können, um daraus eine verlässliche Vorhersage zu machen. Es ist für die Sensoren beispielsweise nicht so ohne weiteres erkennbar, ob die Versuchsperson eine schwierige mathematische Gleichung löst oder einen Einkaufszettel zusammenstellt. Zudem wurde im Gespräch mit den Testteilnehmern klar, dass sie die Testgeräte oftmals am Arbeitsplatz auf dem Tisch gelegen haben und so mancher Sensor nur wenig ändernde Zustände messen konnten, was für die Konstruktion des Entscheidungsbaum hinderlich war.

5.3 Weiterer Verlauf des Experiments

Wie bereits erwähnt, hatte der ursprüngliche Experimentaufbau vorgesehen, dass eine Trainingsphase und danach eine Produktivphase durchgeführt werden würde. Da die Vorhersagen auf der Basis des ersten Teils nur ungenügend waren, wäre es sinnlos gewesen, den zweiten Teil des Experiments trotzdem durchzuführen.

Um zu zeigen, wie das Endresultat ausgesehen hätte, wenn wir mit dem besten Resultat aus der Trainingsphase das Experiment fortgesetzt hätten, haben wir folgende Berechnungen gemacht. Sie basieren auf den KLT4-Werten bei 240 Sekunden Datenaufzeichnung von Proband #3, mit Threshold 8.98 (siehe Tabelle 5.1):

Von den insgesamt **149** Events wären nur gerade **8** als störend klassifiziert worden, welche auch wirklich störend waren (**richtig-positiv**). Dies entspricht **5.3%**. Daneben wären **7** Events als nicht störend eingestuft worden, obwohl sie gestört hätten (**falsch-negativ**). Zusätzlich wären noch **24** Events als störend klassifiziert worden, die dies in Wirklichkeit gar nicht waren (**falsch-positiv**). Von den **32** störenden Events, wären nur gerade **8** als solche erkannt worden. Da beim zweiten

Teil des Experiments anstelle der 150 Events nur noch 50 durchgeführt worden wären, hätten sich auch die Trefferquoten auf etwa **2 bis 3** reduziert. Die Vorhersage hätte demnach nur in diesen wenigen Fällen funktioniert und es wäre kaum möglich gewesen eine Leistungssteigerung zu messen.

KLASSIFIZIERT ALS ->	a	b	WIRKLICHKEIT
	110	7	a = 0 - 8.98
	24	8	b = 8.98 - 18

Tabelle 5.1: Vorhersagen des KLT4 (2-Klassen mit Threshold 8.98) für Proband #3 - 240 Sekunden Sensordaten

6

Einschränkungen und Ausblick

In unserem Experiment haben wir versucht herauszufinden, ob die Arbeitsleistung im Büroalltag erhöht werden kann. Die Resultate aus der Trainingsphase haben gezeigt, dass es mit unserem Versuchsaufbau nicht möglich ist, die Leistungsfähigkeit einer Person zu erfassen, was die Voraussetzung für den zweiten Experimententeil gewesen wäre, die zur Erhöhung geführt hätten. Wegen diese Tatsache, haben wir uns entschieden, dass Experiment in der Trainingphase abbrechen und nicht mehr weiter zu verfolgen.

Der Entscheid, die verwendeten Sensoren einzusetzen wurde gefällt, weil sie in vorgängigen Experimenten gute Resultate erzielt hatten. Da wir nur 5 Versuchsgeräte mit dazugehörigen Sensorboards zur Verfügung hatten und die Zeit kurz bemessen war, konnte das Experiment im ersten und einzigen Durchgang, nur an 5 Personen durchgeführt werden.

Der Experimentaufbau könnte man in zwei Bereichen anpassen, um das Ziel der Arbeitsleistungserhöhung trotzdem weiter zu verfolgen. Zum einen könnte man den KLT durch einen anderen Test austauschen, der in konzentrationsintensiven Situationen bessere Resultate liefert. Da es allerdings keinen standardisierten Leistungstest gibt, der in unserem Experiment ohne weiteres eingesetzt werden könnte, müsste ein eigener entwickelt werden. Man könnte beispielsweise auf dem Display des Testgeräts im Sekundentakt Wörter anzeigen lassen, die danach in der richtigen Reihenfolge notiert werden müssten und die Fehlerquote auf die geistige "Flexibilität" Hinweise geben könnte. Zum anderen wäre die Erweiterung der kontextsensitiven Sensoren. Da die Unterbrechungen in diesem Experiment mehrheitlich im Büroalltag entstanden sind, müsste man darauf spezialisierte Sensoren entwickeln. Es wäre beispielsweise denkbar, dass es eine Kommunikation zwischen dem Tréo und dem PC der Testperson mit Hilfe von Bluetooth geben könnte. Eine Software auf dem PC würde dann durch die Eingaben und dem am Bildschirm angezeigten Inhalt erkennen, ob die Versuchsperson am Arbeiten ist, oder beispielsweise auf einer Webseite die neusten Nachrichten liest. Dieser Arbeitsstatus könnte dann mittels Datenübertragung an den Tréo geschickt werden, wo dieser Faktor in die Vorhersageberechnung des Events einfließen könnte.

Durch den Einsatz von Bluetooth könnte man sich noch eine Vielzahl von zusätzlichen Erweite-

rungen vorstellen. Beispielsweise könnte damit die Lokalisierung stark verbessert werden. Würde das Versuchsgerät erkennen, wenn der Bluetoothadapter der Musikanlage, welche zu Hause steht, im Empfangsbereich auftaucht, kann daraus eindeutig geschlossen werden, dass sich die Testperson zu Hause befindet und dies wäre ein verlässlicher Indikator für den Unterbrechungslevel.

7

Schlussfolgerung

In dieser Arbeit wurde ein Experiment durchgeführt. Dies hatte zum Ziel herauszufinden, ob es möglich ist, mit kleinen und mobilen Sensoren die Arbeitsleistung einer Person zu erhöhen. Obwohl die Durchführung des aufwendigen Experiments als gelungen angesehen werden kann, sind die daraus entstandenen Resultate weniger erfreulich.

Es hat sich gezeigt, dass die berechneten Vorhersagen der Leistungsfähigkeit einer Testperson, mit Hilfe der Sensordaten, nicht von genügender Qualität waren. Der J48-Algorithmus, welcher aus den gesammelten Sensordaten einen Entscheidungsbaum erzeugt und so versuchte die Vorhersagen zu treffen, hat nur ganz vereinzelte Resultate über dem Prior erzeugt. Die Vorhersagen hätten allerdings zwingend über dieser Grenze liegen müssen, um damit eine Optimierung der Arbeitsleistung zu erwirken. In Abbildung 7.1 kann man beispielsweise sehen, dass die pink eingefärbte Linie deutlich unter der blauen Prior-Grenze liegt. Es kann daraus gefolgert werden, dass es mit von uns eingesetzten Sensoren und des Tests, nicht möglich ist, die Arbeitsleistung zu messen und zu erhöhen. Dies hat uns auch veranlasst, das Experiment in der ersten Phase zu beenden und nicht mehr weiter in die Produktivphase zu gehen.

Bei der Auswerten der Resultate wurde weiter noch festgestellt, dass der Unterbrechungslevel nicht mit den Resultaten aus den Rechnungen zusammenhängt.

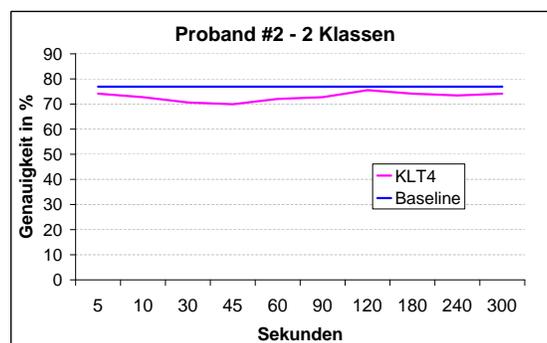


Abbildung 7.1: Vorhersagen der KLT4-Werte für Proband #2

Bei der Suche der Ursache für die schlechten Endergebnisse, konnten zwei Bereiche identifiziert werden, die möglicherweise damit zusammenhängen. Zum Einen können die Resultate durch den durchgeführten KLT verfälscht worden sein. Möglicherweise ist dieser Test nicht in der Lage die Konzentration der Probanden genügend sensibel zu messen. Es müsste demnach ein anderer Test eingesetzt werden, der die Resultate verbessern könnte. Eine andere mögliche Ursache für die schlechten Vorhersagen, könnten die Sensoren gewesen sein. Es kann sein, dass sie nicht genügend sensibel sind, um die Situation zu erfassen, welche die kognitive Leistungsfähigkeit beeinflusst. Um diesem Problem Abhilfe zu schaffen, müssten unter Umständen noch weitere Sensoren dem Experimentaufbau hinzugefügt werden.

A

Anhang

A.1 Antworten aus Probandeninterview

Fragen für Interview mit Probanden (Nach dem Experiment)

1. Kannst du dich rasch auf eine Arbeit konzentrieren?

- #1 Ja
- #2 Nicht speziell
- #3 Ja
- #4 Wenn es wichtig ist, geht es rasch
- #5 Ja

2. Bist du schnell abgelenkt?

- #1 Eher ja
- #2 Ja
- #3 Ja
- #4 Wenn es wichtig ist, geht es rasch
- #5 Kommt auf die Arbeit an. Eher nein

3. Fällt es dir leicht Rechnungen zu lösen ("guter Rechner")?

- #1 Es geht schnell, aber nicht sehr schnell
- #2 Ja

- #3 Ja
- #4 Es geht
- #5 Ja

4. Nach einer Störung dauert es lange bis du wieder in der Tätigkeit drin bist, die du vor der Unterbrechung gemacht hast?

- #1 Nein
- #2 Ja
- #3 Situationsabhängig, eher schnell
- #4 Es ist schwer
- #5 Eher nein

5. Wenn du in einer Situation der starken Konzentration gestört wirst, fällt es dir dann leicht, die gleich starke Konzentration auf eine neue Arbeit zu lenken, die dazwischen kommt?

- #1 Eher Ja
- #2 Nein
- #3 Ja, aber nicht wieder zurück
- #4 Geht gut
- #5 Eher ja

6. Wie hast du das Experiment empfunden? Eher Mühsam, locker? Hattest du eine emotionale Beeinträchtigung (hast du dich schlecht gefühlt, durch das Experiment)?

- #1 Wenn ich unterwegs war, hat es mich am meisten gestört
- #2 Es war nervig, aber nicht besonders gestört
- #3 Eher locker, mitnehmen des Geräts hat gestört
- #4 Unterschiedlich, wenn ich einen schlechten Tag hatte, war es störend.
- #5 War nicht so wild. Das Aufladen des Gerätes war das Mühsamste

7. Hat es dich im täglichen Leben stark gestört, ein solches Gerät mit dir zu nehmen?

- #1 Es ging, war aber nicht schlimm
- #2 Das Sensorboard war ziemlich doof, weil es oft abgeknickt ist
- #3 Ja, es passte nicht sauber in die Hosentasche
- #4 Es war zu gross und die Sensoren waren mühsam. In der Jacke war es auch nicht so toll

- #5 Nein. GPS war mühsam

8. Hast du dich durch mich oder die Technik beobachtet gefühlt?

- #1 Nein
- #2 Nein
- #3 Nein eher nicht
- #4 Kein Problem
- #5 Nein

9. Hattest du Leistungsdruck. Wolltest du nicht, dass die Rechnungsergebnisse schlecht rauskommen? Ehrgeiz?

- #1 Nicht spezieller Leistungsdruck
- #2 Ja, die Aufgaben haben mich angespornt
- #3 Ja, ich wollte gut sein
- #4 Ja
- #5 Geht so. Aber ich habe es ernst genommen und nicht extra schlecht gemacht

10. Hast du immer versucht, möglichst alle Fragen zu beantworten?

- #1 Ja
- #2 Ja
- #3 Ja
- #4 Ja
- #5 Ja

11. Ist dir aufgefallen, dass die Leute in deiner Gegenwart (die vor dem Event mir dir geredet haben) ruhig wurden, wenn du einen Event zu absolvieren hattest?

- #1 Nein
- #2 Eher nicht
- #3 Ganz selten vielleicht
- #4 Eher negative Reaktion
- #5 Ja, auf alle Fälle

12. Glaubst du, dass du mit steigendem Stör-Level schlechter gerechnet hast, oder hattest du das Gefühl, dass du immer etwa gleich performt hast?

- #1 Ja, auf Stufe 5 ging meist recht wenig
- #2 Es war immer ziemlich gleich
- #3 Teilweise, bei hohem Level ging es eher schlecht
- #4 Sehr unterschiedlich, keine Regelmässigkeit
- #5 Bei ganz hohem Level war ich bestimmt schlechter. Ansonsten war es relative gleich

13. Hast du manchmal einen Stör-Level gewählt, der dann beim Rechnen nicht mit deiner Performance übereingestimmt hat? Hättest du also im Nachhinein einen anderen Störlevel gewählt? Eher zu tief/zuhoch gewählt?

- #1 Eher zu tief gewählt, aber nur selten
- #2 Ja
- #3 Teilweise, etwas zu tief
- #4 Ja, manchmal zu tief
- #5 Ja, aber nicht grundsätzlich unterschiedlich (plus/minus 1)

14. Warst du oftmals in gleichen Situationen (Lokalität) während des Experiments?

- #1 Ja, oft am arbeiten am Pult
- #2 Ja
- #3 Ja
- #4 Ja
- #5 Ja fast immer dieselbe Situation. (Mobile auf Arbeitstisch während dem Arbeiten, so ca. 90%)

15. Wie stellst du dir generell ein "Anti-Störungs-System" vor? Eher anhand von Regeln (du sagst, jetzt will ich nicht gestört werden und schaltest ein bestimmtes Profile ein. Es unterliegt alles deiner Kontrolle) oder eher voll automatisch?

- #1 Klare Regeln finde ich besser
- #2 Lieber manuell, ausser es klappt 100%
- #3 Es sollte schon automatisch gehen, das wäre ziemlich cool
- #4 Wäre ziemlich interessant wenn es gehen würde. Vertrauen fehlt
- #5 Manuell würde ich bevorzugen, dann bin ich sicher, dass nicht mal was verloren geht

16. Braucht es überhaupt es überhaupt so etwas?

- #1 Einfache Regeln genügen mir
- #2 Ich bräuchte das nicht unbedingt
- #3 Nice to have
- #4 Wäre schön
- #5 Eher nicht. Ich habe mir angewohnt, das Mobile immer auf stumm zu schalten, wenn eine Störung wirklich unangebracht wäre.

B

Anhang

B.1 Benutzeranweisung für Experiment

Die Benutzeranweisungen wurden allen Experimentteilnehmer abgegeben, um sie auf das Experiment vorzubereiten und den Ablauf zu erklären. Gleichzeitig diente dieses Dokument zum Nachschauen, falls Fehler oder sonstige Ungereimtheiten aufgetaucht sind.



Universität Zürich
Institut für Informatik

Benutzerhinweise für Experiment mit Treo 750 und Sensorboard

1. Grundlegendes zum Experiment und Aufzeichnungsgerät (Treo):

Das Experiment hat zum Ziel herauszufinden, ob es möglich ist die Arbeitsleistung einer Person zu erhöhen, indem die Ablenkung von seiner Arbeit reduziert wird. Anhand von Sensoren sollen Daten erhoben werden, die Hinweise geben ob die Versuchsperson im Augenblick gestört werden sollte oder besser nicht. Um die Konzentration einer Person zu messen, müssen Rechungsaufgaben gelöst werden, die dann ausgewertet werden können.

Um die Daten zu erheben, wird ein Palm Treo eingesetzt, welches mit einer Vielzahl von Sensoren ausgestattet ist. Dazu läuft ein Programm auf dem Treo, welches zu bestimmten Zeitpunkten Sensordaten abspeichert. Während des Experiments sollte das Aufzeichnungsprogramm nicht ausgeschaltet werden. Aus diesem Grund, sollte der Button mit dem roten Hörer (**Siehe Bild I: A**) niemals **im laufenden Experiment** gedrückt werden. Wenn Sie das Experiment in der Nacht unterbrechen möchten, klicken Sie auf die kleine **Exit-Taste** am unteren Rand des Displays und bestätigen Sie mit **JA**. Nun können sie die Taste mit dem roten Hörer (**Siehe Bild I: A**) drücken, um das Gerät ganz auszuschalten.

Am nächsten Morgen müssen Sie am Gerät einen Reset durchführen (**siehe Abschnitt 2**). Bitte öffnen Sie keine Dateien auf dem Gerät, von denen Sie nicht mit Sicherheit wissen, dass sie nicht zur Applikation des Experiments gehören. Geöffnete Dateien verursachen Probleme mit dem Lese/Schreib-Zugriff.

2. Empfohlenes Vorgehen beim Experiment

Um das Experiment schnell abschliessen zu können, ist es wichtig das Gerät so lange wie möglich während des Experiments mit sich mit zu nehmen. Je länger das Gerät an einem Tag arbeitet, desto schneller ist der Sollwert an Daten erreicht, womit das Experiment abgeschlossen ist.

Am besten wird der Treo und GPS-Empfänger am Morgen vom Ladegerät getrennt und das Sensorboard eingesetzt. Um allfällige Systemfehler zu vermeiden, sollte das Gerät nun mit der **Reset-Taste** neu gestartet werden. Sie befindet sich am Gerät rechts unter der Kunststoffklappe (**Siehe Bild III: F**). Falls Sie gerade einen neuen Akku ins Gerät eingesetzt haben, muss die **Reset-Taste** nicht gedrückt werden, da der Treo beim Akkuwechsel bereits neu startet.

Nach dem Start arbeitet das Programm auf dem Treo den ganzen Tag hindurch. Wenn Sie am Abend das Experiment für die Nacht unterbrechen wollen, klicken Sie auf die kleine **Exit-Taste** am unteren Rand des Displays und bestätigen Sie mit **JA**. Jetzt können Sie das Sensorboard abnehmen und das Ladegerät einstecken. Der Treo wird nun geladen und die Datenaufzeichnung ist angehalten. Gleichzeitig sollten Sie auch den GPS-Empfänger ausschalten und neu aufladen. Am darauf folgenden Morgen verfahren Sie gleich, wie bereits oben erwähnt.

Wenn das Gerät während des Experiments anzeigt, dass der Akku keine Leistung mehr hat (Signalton und Anzeige auf dem Display), wechseln Sie am besten den Akku wie unter **Punkt 9 (falls das Gerät noch mit etwas Restenergie läuft)** beschrieben.

Wenn das Gerät aber bereits ausgeschaltet hat, da er Akku **vollständig entladen wurde**, müssen sie zusätzlich zu **Punkt 9** noch **Punkt 10** beachten da das Bluetooth bei vollständigem Entladen des Akkus deaktiviert wurde. Prüfen Sie deshalb bei jedem Start des Geräts, ob das mittlere blaue Licht leuchtet.

Wenn zwei rote Lämpchen (**Siehe Bild II: 1**) beim GPS-Empfänger blinken wird angezeigt, dass dieser aufgeladen werden sollte.

3. Das Lösen der Rechungsaufgaben:

Wenn der Treo ein Event signalisiert hat, muss dieser innerhalb von 30 Sekunden beantwortet werden. Dazu muss zuerst, falls aktiv (siehe **Punkt 8**), die Tastensperre durch Drücken der **seitlichen Taste (Siehe Bild I: E)** deaktiviert werden. Danach ist zu entscheiden wie stark Sie gerade jetzt bei der aktuellen Tätigkeit gestört werden. Geben Sie nun diesen Wert durch Druck auf die entsprechende Bildschirmstaste an (1 für wenig gestört und 5 für massive Störung). Es erscheint nun die Erklärung wie die nachfolgenden Aufgaben gelöst werden sollen. Wenn Sie es wünschen, haben Sie nun 60 Sekunden Zeit um sie zu lesen. Wenn die Aufgaben nicht innerhalb dieser Minute begonnen werden, bricht das Programm ab. Die Aufgabenstellung ist während des ganzen Experiments die gleiche und somit müssen Sie die Aufgabenstellung nicht fortwährend durchlesen.

Wenn nun mit den Aufgaben begonnen wurde, haben Sie **72 Sekunden** Zeit um die Aufgaben zu lösen. Maximal werden 10 nacheinander gestellt. Die 72 Sekunden werden durch einen Fortschrittsbalken angezeigt. Die Resultate können durch drücken der entsprechenden silbernen Zahlen auf der Tastatur (**Siehe Bild I: C**) eingegeben werden und danach muss mit drücken auf das Feld „**Nächste Aufgabe**“ bestätigt werden. Falls bei der Eingabe ein Fehler passiert ist, kann mit der roten „**C-Taste**“ auf dem Bildschirm das Eingabefeld gelöscht werden. Wenn die Zeit abgelaufen ist, schliesst sich das Programm automatisch.

Ziel dieser Aufgaben ist es, Ihre Konzentrationsfähigkeit in der aktuellen Situation zu messen. **Es sollten in der vorgegebenen Zeit möglichst viele Aufgaben richtig gelöst werden.**

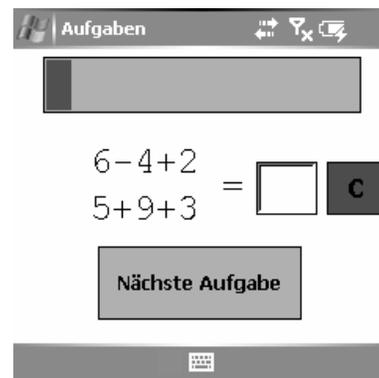
Achtung! Dieser Test misst weder Ihre Intelligenz noch Ihre mathematischen Fähigkeiten und kann auch darüber keine Auskunft geben. Es soll damit nur ermittelt werden, wie leistungsfähig Sie zu dem Zeitpunkt sind und ob Sie den Kopf genügen „frei“ haben um schwierige Rechenaufgabe zu lösen.

Anleitung, wie die Rechnungen gelöst werden sollten:

Zuerst die obere Rechnung ausrechnen und sich das Ergebnis merken. Dann dasselbe mit der Unteren. Beide Resultate im Kopf behalten.

Ist die obere Zahl **grösser** als die Untere, dann muss die Untere von der Oberen **abgezogen** werden. Ist die obere Zahl **kleiner** als die Untere, dann die beiden Zahlen **zusammenzählen**. Es werden so negative Resultate vermieden.

$$\begin{array}{rcl}
 8+9-2 & = & 15 \\
 & (-) & = \mathbf{11} \\
 5-4+3 & = & 4 \\
 \text{*****} & & \\
 3+6-8 & = & 1 \\
 & (+) & = \mathbf{18} \\
 9+1+7 & = & 17
 \end{array}$$



4. Sensorboard

Anbringen des Sensorboards:

Um das Sensorboard am Treo zu befestigen, muss das Board mit den grossen Sensoren nach oben am unteren Ende des Treos eingesteckt werden. Es muss darauf geachtet werden, dass die Kontakte nicht abgedrückt werden und das Board möglichst gerade in die Öffnung am Treo eingesteckt wird.

Entfernen des Sensorboards:

Beim Entfernen muss das Sensorboard gut festgehalten und gerade aus dem Treo gezogen werden. Es ist darauf zu achten, dass das Board nicht herausgeknickt wird. Durch die Hebelwirkung können die Kontakte beschädigt werden.

5. GPS-Empfänger:**Generelles:**

Um den GPS-Empfänger ein/aus zu schalten, muss der **seitliche Knopf am GPS-Empfänger (Siehe Bild II: 1)** etwa 4 Sekunden lang gedrückt gehalten werden bis ein blaues Licht (**Siehe Bild II: 2**) zu blinken beginnt.

Das rote Licht (**Siehe Bild II: 2**) zeigt an, ob der Empfänger genügen GPS-Satelliten-Empfang hat. Wenn es blinkt, ist kein oder nur ungenügender Empfang vorhanden.

Empfang:

Wenn möglich, positionieren sie den GPS-Empfänger in der Nähe eines Fensters um bestmöglichen Empfang zu gewährleisten (rotes Licht leuchtet konstant). Es kann vorkommen, dass der Empfänger auch in Fensternähe keine Verbindung aufbaut. In diesem Fall können Sie beispielsweise den Empfänger einen kurzen Moment auf das äussere Fenstersims legen bis die Verbindung steht (Er braucht freie Sicht, vertikal nach oben für ca. 30 Sekunden). Danach sollte der Empfang auch im Gebäude in der Nähe der Fenster gewährleistet sein.

Die Verbindung vom Treo zum GPS-Empfänger funktioniert über Bluetooth. Am besten tragen Sie den GPS-Empfänger mit sich mit oder sie entfernen sich nicht weiter als 10m.

Akku des GPS-Empfängers:

Wenn beim Empfänger zwei rote Lichter (**Siehe Bild II: 2**) zu sehen sind, muss der Empfänger aufgeladen werden. Der Ladevorgang ist beendet, wenn das grüne Ladelicht erloschen ist. Falls der GPS-Empfänger während des Tages aufgeladen werden muss, kann dies auch bei eingeschaltetem Gerät erfolgen und das Experiment muss nicht abgebrochen werden.

6. Docking Station:

Die Dockingstation kann für das Aufladen des zweiten Akkus verwendet werden. Der zweite Akku sollte die Zeit überbrücken, in der Sie keine Gelegenheit haben das Gerät am Netz aufzuladen. Um einen Akku aufzuladen, öffnen sie die Klappe auf der Vorderseite der Dockingstation und legen Sie den Akku so ein, dass die Kontakte anliegen. Vergewissern Sie sich, dass das rote Lämpchen auf der Vorderseite der Dockingstation zu leuchten beginnt. Der Akku ist vollständig geladen, wenn das Lämpchen grün leuchtet.

7. Aufladen:

Um das Gerät aufzuladen, klicken Sie auf die kleine **Exit-Taste** am unteren Rand des Displays und bestätigen Sie mit **JA**. Nun können Sie das Sensorboard entfernen und das Ladegerät auf der Unterseite rechts einstecken (Der kleine weisse Pfeil auf dem Ladestecker muss nach oben zeigen).

Der Akku ist vollständig geladen, wenn am Gerät oben links ein grüne Lämpchen (**Siehe Bild I: D**) leuchtet. Sie können den Ladestatus auch durch klicken auf das Akkusymbol auf dem Bildschirm oben rechts überprüfen.

Nach dem Aufladen können Sie das Ladekabel entfernen, das Sensorboard wieder einsetzen und durch drücken der **Reset-Taste** das Experiment weiterführen. Sie befindet sich am Gerät rechts unter der Kunststoffklappe (**Siehe Bild III: F**).

8. Bildschirmschoner/Tastensperre aktivieren/deaktivieren:

Um zu vermeiden, dass ungewollt Tasten gedrückt oder Kommandos auf dem Bildschirm eingegeben werden, besitzt der Treo eine Tasten-Sperr-Funktion. Um den Ablauf des Experiments nicht unnötig zu stören, empfiehlt es sich, diese Tasten-Sperre immer aktiviert zu haben, solange keine Eingaben am Gerät zu machen sind. Für die Aktivierung wie auch zur Deaktivierung muss auf der linken Geräteseite die **kleine alleinstehende Taste (Siehe Bild I: E)** etwa 2 Sekunden lang gedrückt werden. Die Tasten-Sperre wird **nicht** automatisch nach einer gewissen Zeit eingeschaltet und muss deshalb immer von Hand aktiviert werden.

9. Akku wechseln:

Der Akku kann während des laufenden Experiments ausgetauscht werden. Gehen Sie wie folgt vor:

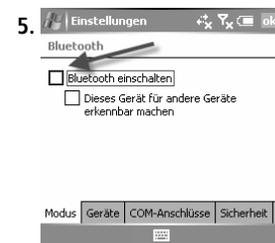
- ✓ Entfernen Sie die Akkuabdeckung indem Sie den eingelassenen Knopf (**Siehe Bild IV: 1**) auf der Rückseite des Treos drücken und den Deckel der Länge nach weg verschieben.
- ✓ Den Akku anheben und durch den neuen austauschen.
- ✓ Das Gerät startet nun automatisch.
- ✓ **ACHTUNG!** Wenn der Akku vor dem Austauschen des Akkus **bereits ausgeschaltet hat**, da der Akku völlig entladen war zusätzlich **Punkt 10** durchführen

10. Restart nach vollständiger Entladung des Akkus:

Setzen Sie neuen Akku ein oder laden Sie das Gerät vollständig auf. Wenn der Akku völlig entleert hat (Gerät kann nicht mehr eingeschaltet werden), muss nach dem Systemstart die Bluetooth-Verbindung manuell eingeschaltet werden. Bitte warten Sie mit dem unterstehenden Vorgehen bis das Gerät vollständig gestartet ist und das Aufzeichnungsprogramm auf dem Display erscheint.

Vorgehen:

1. **Windows-Taste**  auf der Tastatur oben links drücken (**Siehe Bild I: B**)
2. Menüpunkt **Einstellungen** auswählen
3. Im Register am unteren Bildrand **Verbindungen** auswählen
4. Auf **Bluetooth** klicken
5. Bei Bluetooth einschalten einen **Haken setzen**
6. Oben rechts auf **OK** drücken
7. Fertig



11. Fehlermeldungen und Vorgehen:

Fehlersignal ist zu hören, Treo vibriert und es wird angezeigt, dass die Sensorboard-Verbindung nicht funktioniert:

- Wenn Sensorboard nicht eingesteckt ist, stecken Sie es bitte ein.
- Wenn Sensorboard bereits eingesteckt ist bitte kurz aus- und wieder einstecken (ca. 1 Sekunde warten).

Auf dem Bildschirm erscheint ein Fenster mit einer Fehlermeldung:

Beenden anklicken und Treo mit Reset-Taste neu starten. Dazu muss die kleine Kunststoffklappe auf der rechten Geräteseite geöffnet werden und die kleine Drucktaste mit einem Stift eingedrückt werden. **(Siehe Bild III: F)** Der Treo startet sich neu.

Beispiel einer Fehlermeldung:



Der Treo kann nicht mehr eingeschaltet werden (Schwarzer Bildschirm und keine Tastaturbeleuchtung bei Druck auf die Tasten. Wenn Tasten noch leuchten siehe Punkt 4. Bildschirmschoner/Tastensperre aktivieren/deaktivieren):

Die Batterie ist vollständig entladen. Bitte lesen weiter unter **Punkt 10** Restart nach vollständiger Entladung des Akkus.

Rote Signalleuchte über dem PALM Logo blinkt und Akku-Warnhinweis wird angezeigt:

Der Akku ist leer. Bitte Laden Sie das Gerät auf wie unter **Punkt 7** beschrieben auf oder tauschen Sie den Akku aus wie in **Punkt 9** erklärt.

Zwei rote Lichter blinken am GPS-Empfänger:

Der Akku des GPS-Empfängers muss aufgeladen werden.

Das mittlere blaue Licht blinkt am GPS-Empfänger nachdem das Gerät neu gestartet wurde:

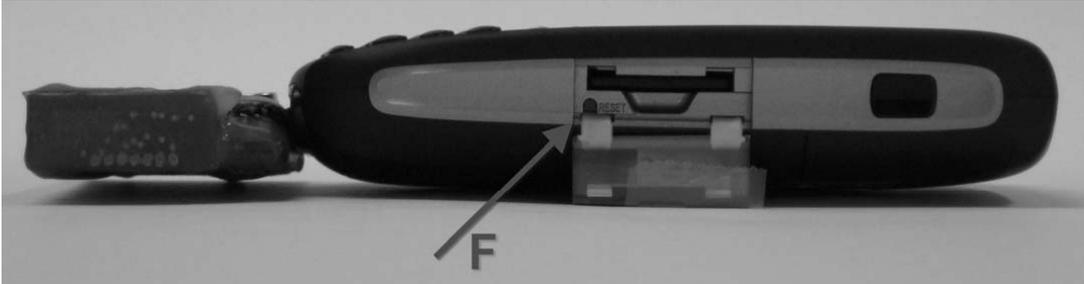
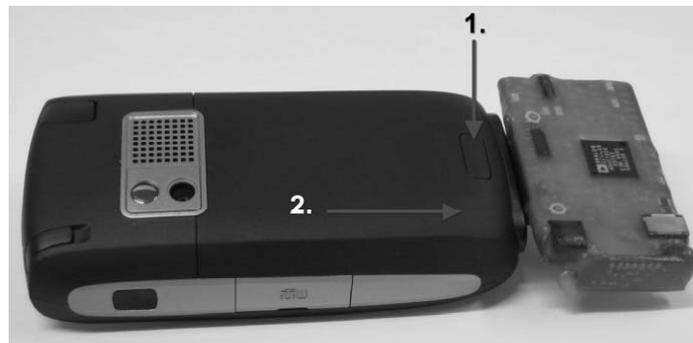
Wenn das blaue Licht ca. 2 Minuten nach dem Neustart des Treos immer noch blinkt, überprüfen Sie bitte ob die Bluetooth-Verbindung vom Treo zum GPS-Empfänger aktiviert ist. Verfahren Sie bitte wie unter **Punkt 10** beschrieben und prüfen Sie, ob der Hacken an entsprechender Stelle gesetzt ist. Ist dies der Fall, ist alles in Ordnung.

Bild I:



Bild II:



Bild III:**Bild IV:**

C

CD

Dieser Arbeit ist eine CD-Rom beigelegt. Sie beinhaltet diese Arbeit in elektronischer Form (als PDF), die Resultate des durchgeführten Experiments, sowie sämtlichen Quellcode der in dieser Arbeit vorgestellten Software.

Die Tabellen C.1 zeigen eine Übersicht über den Inhalt des Datenträgers.

Verzeichnis	Inhalt
Diplomarbeit	Diplomarbeit als PDF
Zusammenfassung	Zusammenfassung (Deutsch & Englisch) der Diplomarbeit
Resultate	Resultate des Experiments
LaTeX	LaTeX- Dateien dieser Diplomarbeit
Software	Quellcode der Software

Tabelle C.1: Inhalt CD-ROM

Abbildungsverzeichnis

2.1	Unterbrechungsfaktoren [34]	6
2.2	Hierarchische Aufteilung einer Arbeit [1]	7
3.1	Phasen und Bestandteile des Experiments	12
3.2	Beispiel von zwei KLT-Aufgaben [9]	15
3.3	Beispiel einer d2-Aufgaben	16
3.4	Beispiel der FAIR-Zielitems	18
3.5	Das Experimentiergerät - Der Tréo 750	22
3.6	Das Sensorboard	23
3.7	Datensatz des Sensorboards	24
3.8	GUI der CADT-Applikation	26
3.9	Darstellung des KLT auf dem Display des Treo 750	28
3.10	Ablauf der KLT-Applikation	29
3.11	Auswahl des Unterbrechungslevel auf dem Tréo	34
3.12	Darstellung der KLT-Anleitung auf dem Tréo	34
4.1	Schematischer Ablauf der Datenaufbereitung und Analyse	39
4.2	Verteilung der Events der Probanden #1 und #2	41
4.3	Unterbrechungslevel Proband #1	42
4.4	Unterbrechungslevel Proband #2	42
4.5	Eventverteilung aller Probanden	43
4.6	Korrelation Unterbrechungslevel und KLT1 - #1	45
4.7	Korrelation Unterbrechungslevel und KLT4 - #1	45
4.8	Korrelation Unterbrechungslevel und KLT5 - #1	46
4.9	Korrelation Unterbrechungslevel und KLT1 - #3	46
4.10	Korrelation Unterbrechungslevel und KLT4 - #3	47
4.11	Korrelation Unterbrechungslevel und KLT5 - #3	47
4.12	Häufigkeiten der KLT1- & KLT4 - #1	49
4.13	Häufigkeiten der KLT1- & KLT4 - #2	49

4.14	Häufigkeiten der KLT1- & KLT4 - #3	49
4.15	Vorhersagen für Proband #1	50
4.16	Vorhersagen für Proband #2	51
4.17	Vorhersagen für Proband #3	51
4.18	Vorhersagen der Unterbrechungslevel (5-Klassen) für Proband #1	52
4.19	Vorhersagen der Unterbrechungslevel (5-Klassen) für Proband #2	53
4.20	Vorhersagen der Unterbrechungslevel (5-Klassen) für Proband #3	53
4.21	Vorhersagen der KLT1-Werte (5-Klassen) für Proband #1	54
4.22	Vorhersagen der KLT1-Werte (5-Klassen) für Proband #2	54
4.23	Vorhersagen der KLT1-Werte (5-Klassen) für Proband #3	55
7.1	Vorhersagen der KLT4-Werte für Proband #2	65

Tabellenverzeichnis

3.1	Entscheidungsmatrix für Konzentrationstest (3 = beste Note)	20
3.2	Verwendete Software von Drittherstellern	25
3.3	Anpassungen CADT-Code	27
3.4	Angaben zu den Probanden	32
4.1	Verteilung aller Events auf die einzelnen Unterbrechungslevel	44
4.2	Vorhersagen des KLT4 - #1	50
4.3	Vorhersagen des KLT1 - #1	50
4.4	Vorhersagen Unterbrechungslevel - #2	52
4.5	Vorhersagen KLT1 - #1	53
5.1	Vorhersagen des KLT4 - #3	61
C.1	Inhalt CD-ROM	81

Literaturverzeichnis

- [1] ADAMCZYK, Piotr D.; BAILEY, Brian P.: If not now, when?: the effects of interruption at different moments within task execution. In: *CHI '04: Proceedings of the SIGCHI conference on Human factors in computing systems*. New York, NY, USA : ACM, 2004. – ISBN 1-58113-702-8, S. 271-278
- [2] AMELANG, Manfred: *Psychologische Diagnostik und Intervention*. Springer, 2002
- [3] BENZ, Domenic: *Voraussage von Benutzerverhalten in dynamischen Umgebungen*, Universität Zürich, Diplomarbeit, 2007
- [4] CHRISTIANI, Stefan: *A study on activity and location recognition using various sensors*. Universität Zürich, Institut für Informatik, Bachelorarbeit, 2007
- [5] CUTRELL, Czerwinski M. E.; HORVITZ, E.: Notification, disruption, and memory: Effects of messaging interruptions on memory and performance. In: *INTERACT '01 (2001)*, S. 263-269
- [6] CZERWINSKI, Mary; CUTRELL, Edward; HORVITZ, Eric: Instant Messaging and Interruption: Influence of Task Type on Performance. In: *Microsoft Research (2000)*
- [7] DABBISH, Laura; KRAUT, Robert: Coordinating communication: awareness displays and interruption. In: *CHI '03: CHI '03 extended abstracts on Human factors in computing systems*. New York, NY, USA : ACM, 2003. – ISBN 1-58113-637-4, S. 786-787
- [8] DETWEILER, Hess S. M. M. C. C. M. C.; PHELPS, M. P.: *Interruptions and working memory. Technical report, The Pennsylvania State University, Department of Psychology*. 1994
- [9] DÜKER, Heinrich: *K-L-T Der Konzentrations-Leistungs-Test*. Verlag für Psychologie, 1965
- [10] E. CUTRELL, M. C.; HORVITZ, E.: Notification, Disruption, and Memory: Effects of Messaging Interruptions on Memory and Performance. In: *Proceedings of Interact 2001 (2001)*
- [11] EBBINGHAUS, H.: *Memory: A contribution to experimental psychology*. 1885
- [12] EGGER, P.: *Unterbrechbarkeitsbestimmung für wearable computing applikation*, Universität Zürich, Institut für Informatik, Diplomarbeit, 2004

- [27] MCFARLANE, Daniel C.: Interruption of People in Human-Computer Interaction. In: *INTERACT'99* (1999), S. 295–303
- [28] NORMAN, Y. Miyata; D.: Psychological issues in support of multiple activities. In: *D. A. Norman and S. W. Draper (Eds.), User Centered Systems Design: New Perspectives on Human-Computer Interaction*. (1986), S. 265–284
- [29] QUINLAN, J. R.: *C4.5: programs for machine learning*. San Francisco, CA, USA : Morgan Kaufmann Publishers Inc., 1993. – ISBN 1–55860–238–0
- [30] REICHEL, Wolfgang: *Der grosse Intelligenztest*. Klett-Cotta, 2006
- [31] SCHILLER, Jochen: *Mobile communications*. Boston, MA, USA : Addison-Wesley Longman Publishing Co., Inc., 2000. – ISBN 0–201–39836–2
- [32] SCHMIDT, Albrecht; TAKALUOMA, Antti; MÄNTYJÄRVI, Jani: Context-Aware Telephony Over WAP. In: *Personal Technologies 4-4* (2000), S. 225–229
- [33] SMITH, Allen E. Milewski ; Thomas M.: Providing presence cues to telephone users. In: *CSCW '00: Proceedings of the 2000 ACM conference on Computer supported cooperative work*. New York, NY, USA : ACM, 2000. – ISBN 1–58113–222–0, S. 89–96
- [34] STOUFFS, Alexandre: Interruptions as Multimodal Outputs: Which are the Less Disruptive? In: *ICMI '02: Proceedings of the 4th IEEE International Conference on Multimodal Interfaces*. Washington, DC, USA : IEEE Computer Society, 2002. – ISBN 0–7695–1834–6, S. 479
- [35] VORBURGER, Peter; BERNSTEIN, Abraham: Towards an Artificial Receptionist: Anticipating a Persons Phone Behavior / University of Zurich, Department of Informatics.
Abrufdatum: 2005. http://www.ifi.uzh.ch/ddis/staff/goehring/btw/files/tech_vorburger_large.pdf. 2005
- [36] WEISE, Georg: *Psychologische Leistungstests Band 1*. Verlag für Psychologie, 1975