Combining Physiological Data and Context Information as an Input for Mobile Applications

Dissertation
zur Erlangung des Doktorgrades
der Naturwissenschaften

vorgelegt beim Fachbereich 12
der Johann Wolfgang Goethe-Universität
in Frankfurt am Main

von
Claudia Stockhausen
aus Bad Soden-Salmünster

Frankfurt 2016
(D 30)

vom Fachbereich Informatik und Mathematik (12) der

Johann Wolfang Goethe-Universität als Dissertation angenommen.

Dekan:	Prof. Dr. Uwe Brinkschulte
Gutachter:	Prof. Dr. Detlef Krömker
	Prof. Dr. Wolfgang Müller
	Prof. Dr. Ralf Dörner

Datum der Disputation:

Acknowledgments

During this thesis, i have been supported by several people during the last years.
Foremost, i would like to thank my advisor Prof. Dr. Detlef Krömker, who gave me the opportunity and support to write this thesis. Besides my advisor, i would as well like to thank committee members Ralf Dörner and Wolfgang Müller for their helpful feedback and interesting discussions.
As a huge part of this thesis was carried out during my time working as a research assistant at Goethe Universtät Frankfurt, i also want to take the opportunity to express my gratitude to my colleagues at GDV and studiumdigitale for the last few years of support, discussions and fun.
Finally i would like to thank my husband Sven Walter, who supported and motivated me through all these long days and nights of work.

Zusammenfassung

Moderne Smartphones stellen eine Vielzahl an Möglichkeiten zur Erfassung von Daten zur Verfügung. Sensoren zur Positionsbestimmung können Aufschluss über den Ort eines Nutzers geben. Kontextinformationen aus Anwendungen bieten ein Bild über die aktuelle Leistung des Nutzers. Eine Vielzahl an weitern Sensoren bieten zudem die Möglichkeit physiologische Daten eines Nutzers zu messen. Von einer Einschätzung des Benutzerzustands können Anwendungen profitieren, in dem Benutzerinformationen oder Funktionalitäten auf den Zustand zugeschnitten werden können. Ein Benutzer unter Stress profitiert gegebenenfalls von anderen Aspekten als ein Benutzer, dessen Zustand als gelangweilt bewertet wird.
Es ist möglich mittels entsprechender Sensoren eine Vielzahl physiologischer Signale die vom menschlichen Körper erzeugt werden zu erfassen. Insbesonders Signale wie Herzrate oder elektrodermale Aktivität können genutzt werden, um Rückschlüsse auf den aktuellen Zustand des Benutzers zu ziehen. Die Herzrate eines Benutzers kann zum Beispiel nicht nur dafür genutzt werden um zu bestimmen wie aufgeregt dieser ist, sondern erlaubt mittels weiterführender Verarbeitung der entsprechenden Messwerte auch eine Bestimmung der mentalen Belastung des Nutzers. Andere Maße wie elektrodermale Aktivität lassen einen Rückschluss auf die emotionale Aufregung eines Benutzers zu.

Sowohl Herzrate als auch elektrodermale Aktivität können zuverlässig über relativ kleine und drahtlose Sensoren gemessen und an mobile Geräte wie Smartphones übertragen werden. Zur Messung der Herzrate bieten sich Sensoren an wie sie beispielsweise im Sport genutzt werden, da sie eine ausreichende Genauigkeit bieten und weit verbreitet sind. Elektrodermale Aktivität benötigt zwei Sensoren auf der Haut an bestimmten Stellen des Körpers. Hierfür gibt es Lösungen, die ähnlich wie ein Armband getragen werden können. Kontextinformationen, die Rückschlüsse auf Position und Bewegung zulassen, können über eingebaute Sensoren eines mobilen Gerätes wie GPS oder Accelerometer gemessen werden.

Es existieren verschiedene Modelle um den Zustand einer Person näher zu beschreiben. Russell veranschaulicht in einem zweidimensionalem Modell mit Hilfe der Achsen Aufregung und Valenz verschiedene Emotionen, wie z.B. Stress, Langweile oder Entspannung. Yerkes und Dodson beschreiben den Zusammenhang zwischen Aufregung und Performanz. In der von Yerkes und Dodson beschriebenen Kurve wird je nach Schwierigkeit der Aufgabe eine hohe Performanz bei einem mittlerem Aufregungslevel erreicht. Zu hohe oder zu niedrige Aufregungslevel führen zu einer niedrigeren Performanz.

Es existieren zwar einige Arbeiten und Modelle, die auf Basis physiologischer Daten den Benutzerzustand bestimmen, jedoch eignen sich hiervon nur wenige um in mobilen Szenarien genutzt zu werden. Die Kombination mit Kontextinformationen wurde bisher

ebenfalls nur eingeschränkt untersucht.

In dieser Arbeit wird ein Modell vorgestellt, welches Kontextinformationen und physiologische Daten eines Nutzers als Eingabe für mobile Anwendungen nutzt. Auf Basis der Daten wird der aktuelle Benutzerzustand bestimmt, welcher abhängig von der Art der Anwendung für Adaptionen genutzt werden kann. Mobile Szenarien stellen dabei verschiedene Anforderungen an das Modell. Zum einen sind physiologische Daten teilweise durch Bewegung beeinflussbar, zum anderen befindet sich der Nutzer nicht in einer kontrollierten Umgebung und ist verschiedenen ablenkenden Faktoren ausgesetzt. Auch muss in mobilen Szenarien berücksichtigt werden, dass durch unvorhergesehene Ereignisse wie den Ausfall eines Sensors auch zu Interpretation notwendige Daten wegfallen können.

Um einem möglichst breitem Spektrum an mobilen Anwendungen gerecht zu werden, bietet das Modell sowohl emotionale als auch kognitive Aspekte als Ausgabe. Der emotionale Zustand des Benutzers basiert dabei auf dem zweidimensionalen Ansatz von Russell um verschiedene Zustände zu unterscheiden. Um die kognitiven Aspekte abzudecken, wird die mentale Belastung des Benutzers bestimmt.

Für das Modell werden elektrodermale Aktivität und Herzrate als physiologische Signale genutzt, da beide Signale mit relativ kleinen Sensoren messbar sind welche den Benutzer nicht behindern. Als Kontextinformationen werden Bewegung und Performanz genutzt. Schritte werden über Accelerometer erfasst, welche in einer Vielzahl mobiler Geräte integriert sind. Die Performanz eines Benutzers wird von der Anwendung selbst an das Modell übermittelt.

Das eigentliche Modell basiert auf einem Fuzzy Logik basiertem Ansatz. Fuzzy Logik ermöglicht es Unschärfe auszudrücken sowie einen kontinuierlichen Strom an Daten zu verarbeiten. Ein weiterer Vorteil des Fuzzy Logik Ansatzes sind die relativ einfach konfigurierbaren Regeln, für die keine tieferen Kenntnisse im Programmcode notwendig sind da sie in nahezu normaler Sprache formuliert werden können.

Vor Nutzung der Daten als Eingabe im Modell werden verschiedene Schritte zur Vorverarbeitung der Daten vorgenommen. Danach werden die Eingangsdaten fuzzifiziert und ihnen basierend auf der jeweiligen Fuzzy Funktion Werte zugewiesen. Eingangsdaten im Modell sind Herzrate, Herzraten Variabilität, elektrodermale Aktivität, Performanz des Benutzers und Anzahl der Schritte. Schritte werden in drei (niedrig, mittel und hoch) und Performanz in fünf Klassen (sehr niedrig bis sehr hoch) unterteilt. Die physiologischen Signale werden jeweils in fünf Klassen (niedrig bis hoch) unterteilt.

In zwei Schritten werden die Eingabesignale in emotionale und kognitive Werte umgewandelt. Über ein erstes Set bestehend aus mehreren Fuzzy Regeln werden die Eingabesignale in Aufregung, Valenz, mentale Belastung und Kontext umgewandelt. Für Aufregung und Valenz sind sieben verschiedene Klassen möglich (sehr niedrig bis sehr hoch).

In einem zweiten Schritt werden die Ergebnisse aus dem ersten Schritt, Aufregung, Valenz und Kontext, anhand von weiteren Fuzzy Regeln in die emotionalen Benutzerzustände umgewandelt. Mit den Werten von Valenz und Aufregung wird ein 7x7 Gitter basierend auf dem Modell von Russell genutzt um Werte von acht emotionalen Zuständen zu bestimmen.

Kontextinformationen wurden im Rahmen von Performanz in die Fuzzy Regeln zur Überführung in die acht emotionalen Zustände integriert. Diese acht Zustände, arlarmiert, aufgeregt, glücklich, entspannt, müde, gelangweilt, traugrig und frustriert, haben jeweils vier Klassen, die ihren Wert beschreiben (sehr niedrig, niedrig, mittel und hoch). Die acht emotionalen Zustände mit ihren jeweiligen Werten sind die Ausgabe für Anwendungen, die den emotionalen Zustand nutzen.

Der kognitive Benutzerzustand wird aus Performanz des Benutzers, Anzahl der Schritte und mentaler Belastung überführt. Der kognitive Zustand wird in vier Klassen unterteilt: niedrig, mittel, hoch und sehr hoch. Da gerade hohe kognitive Belastungen interessant sind, wurde der hohe Bereich in hoch und sehr hoch unterteilt.

Die Werte werden am Ende über ein Verfahren zur Defuzzifierung wieder in Werte umgerechnet. In dieser Arbeit wurde das Verfahren der gewichteten Mittelwerte genutzt. Die Werte werden an die jeweiligen nutzenden Anwendungen weitergegeben, welche basierend auf den Ausgabewerten des Modelles mögliche Adaptionen bestimmen und anwenden.

Das Modell wurde als ein im Hintergrund laufender Service für das Betriebssystem Android implementiert. Zur Konfiguration der Sensoren wird ein simples Userinterface geboten. Vor Nutzung des Modells wird eine cirka fünf minütige Baseline Messung empfohlen. Der Service stellt die Ergebnisse des Modells über eine Schnittstelle anderen Anwendungen zur Verfügung. Während der Laufzeit des Services werden aktuelle Werte im User Interface angezeigt.

Verschiedene Anwendungen wurden entwickelt und genutzt um das entwickelte Modell zu evaluieren. Die Anwendungen decken dabei unterschiedliche Arten von Anwendungstypen ab, um verschiedene Aspekte des Modells zu evaluieren. Anwendungen zur Unterhaltung könnten mehr vom emotionalen Zustand profitieren, wohingegen leistungsorientierte Anwendungen vom kognitiven Zustand profitieren können.

Das Spiel "Zone of Impulse" ist ein Weltraum-Shooter und passt verschiedene Spielelemente basierend auf dem Benuterzustand an, um einen Ausgleich zwischen zwei gegeneinander spielende Benutzer zu schaffen. Angepasst werden unter anderem Elemente wie die Aufladezeit einer Spezialfähigkeit oder die Geschwindigkeit des eigenen Raumschiffes. Das Spiel nutzt vor allem den emotionalen Zustand eines Benutzers. Zunächst wurde bestimmt, welcher der acht Ausgabezustände erwünscht sind und welche nicht. Die beiden Zustände in denen der Nutzer aufgeregt oder glücklich sind werden dabei angestrebt. Zwei weitere Zustände wurden als Übergangszonen definiert und vier Zustände als nicht erwünscht. Abhängig davon, ob ein Benutzer in einem erwünschtem, einem Übergangszustand oder einem unerwünschten Zustand war, wurde keine Adaption, eine einfache Adaption oder eine erhöhte Adaption durchgeführt, um den Benutzer in einen erwünschten Zustand zu führen oder zu halten.

Eine weitere genutzte Anwendung zur Evaluation ist ein Vokabeltrainer. Die Vokabeltrainer Applikation bietet zu einer Vokabel jeweils vier mögliche Antworten an. Innerhalb eines Zeitlimits muss die richtige aus den vier Vorgaben ausgewählt werden um Punkte zu erreichen. Sowohl emotionaler als auch kognitiver Zustand wurden zur Adaption der Schwierigkeit genutzt. Der emotionale Zustand wurde zunächst in zwei erwünschte und drei unerwünschte Zustände eingeteilt. Drei weitere Zustände sind als Übergangszustände definiert. Für den kognitiven Zustand war der mittlere Bereich erwünscht, ein hoher Wert wurde als Übergangszustand gesehen. Niedrige und sehr hohe Werte für den kognitiven Zustand waren unerwünscht.

Zur Evaluierung des Modells wurden weitere Anwendungen untersucht. Unter anderem wurde eine Anwendung mit Informationen rund um einen Flughafen, die basierend auf dem Benutzerzustand eines von drei verschiedenen User Interfaces anezeigt näher untersucht. Gestresste Benutzer erhalten ein reduziertes Interface, wohingegen gelangweilte Benutzer mehr Funktionalitäten zur Verfügung gestellt bekommen.

Außerdem wurden erste Untersuchungen zur Kombination von Benutzerzustand und der Wahl des Level of Detail beim Rendern von Videos durchgeführt. Weitere genutzte Anwendungen waren unter anderem ein Spiel, welches physiologische Daten als direkte Eingabe zur Steuerung des Spiels nutzt sowie ein Adaptionsmanger, welcher basierend auf dem Benutzerzustand verschiedene Einstellungen des Telefons änderte, wie z.B. den Klingelton in stressigen Situationen lautlos zu stellen.

Im Rahmen einer Studie wurden verschiedene Aspekte des Modells mit Hilfe von zwei Anwendungen evaluiert. Um sowohl kognitive als auch emotionale Aspekte abzudecken, wurden das Spiel ”Zone of Impulse” und der Vokabeltrainer zur Evaluierung herangezogen. In der Studie wurden beide Anwendungen mit Modell, ohne Modell und mit einem teilweise integriertem Modell in verschiedenen Szenarien gegenüber gestellt. Ein Teil der insgesamt 41 Studienteilnehmer waren in einer Versuchsgruppe, die den Test außerhalb der kontrollierten Laborumgebung auf der Straße durchführte. Verschiedene Bewertungen zu Spaß, Überforderung und Unterstützung wurden erfasst sowie zusätzlich die kognitive Belastung über den NASA-TLX Fragebogen.

Die Ergebnisse der Studie unterstützen die These, dass die Kombination aus physiologischen Daten und Kontextinformationen die Interpretationsqualität des Benutzerzustandes verbessern. In der Studie wurde für ”Zone of Impulse” eine Version mit komplettem Modell und eine Version ohne Kontextinformationen gegenüber gestellt. Spaßwurde in der Version mit intgrierten Kontextinformationen signifikant besser bewertet. Der Aspekt Überforderung wurde nicht signifikant besser bewertet.

Die Verbesserung der Interpretationsqualität, wenn physiologische Daten und Kontextinformationen kombiniert werden, wird auch von einem Vergleich zwischen einer Version mit und einer Version ohne Modell für beide Anwendungen unterstützt. Für beide Anwendungen wurde die Version mit intergriertem Modell in den NASA-TLX Fragebögen signifikant besser bewertet. In den subjektiven Fragebögen wurden die meisten Aspekte nicht signifikant besser bewertet.

Neben der Verbesserung der Interpretation durch Kontextinformationen wurde auch untersucht, ob das Modell in mobilen Szenarien zuverlässige Ergebnisse liefert und akzeptabel ist. Der Tragekomfort der Sensoren wurde von den Versuchsteilnehmern gut bewertet, sie fühlten sich durch die Sensoren nicht eingeschränkt. Neben dem Tragekomfort wurden beide Anwendungen in unterschiedlichen Szenarien miteinander verglichen. Ein Teil der Versuchsteilnehmer befand sich in einer sitzenden Position in einem Gebäude. Der andere Teil der Versuchsteilnehmer lief während des Tests außerhalb des Gebäudes umher und musste teilweise während der Nutzung der mobilen Anwendungen mit der Umwelt interagieren. Die Ergebnisse zeigten, dass es zwischen beiden Testgruppen keine signifikanten Unterschiede in der Bewertung der verschiedenen Aspekte gibt. Das führt zu der Vermutung, das die Zuverlässigkeit der Interpretation in mobilen Szenarien nicht von stationären Szenarien abweicht.

Ein weiterer wichtiger Aspekt ist die Robustheit des Modelles, da in mobilen Szenarien die Möglichkeit besteht, dass einer der Datenkanäle ausfällt. In der Studie wurde für beide Anwendungen jeweils eine Version erstellt, in der ein Eingabesignal der physiologischen Daten entfernt wurde. Die Ergebnisse der verschiedenen Fragebögen zeigten, dass zwischen beiden Versionen kein signifikanter Unterschied besteht und der Verlust eines Eingabesignals kompensiert werden konnte. Jedoch wurden in der Studie relativ normale Situationen getestet. In extremeren Situationen besteht die Möglichkeit, dass der Verlust eines Eingabesignals sich deutlicher bemerkbar macht.

Die Interpretation des Benutzerzustands auf Basis verschiedener Informationsquellen ist ein bedeutsames Thema in der Forschung. Es existieren Modelle und Anwendungen, die physiologische Daten zur Interpretation des Benutzerzustandes heranziehen. Bisher wurde dies jedoch nur eingeschränkt in Bezug auf mobile Anwendungen und der Nutzung in mobilen Szenarien betrachtet. In dieser Arbeit wurde ein Modell vorgestellt, dass durch die Kombination physiologischer Daten mit Kontextinformation versucht die Herausforderungen, die in mobilen Szenarien auftreten, zu addressieren. Eine durchgeführte Studie zeigte positive Ergebnisse für verschiedene evaluierte Aspekte wie Robustheit, Nutzerakzeptanz der Sensoren sowie Zuverlässigkeit der Interpretationsergebnisse in mobilen Szenarien.

Das vorgestellte Modell wurde für einen kleinen Teil an möglichen Situationen evaluiert. Weitere Untersuchungen mit anderen Anwendungen, anderen Situationen oder anderen physiologischen Daten sind notwendig, um ein genaueres Urteil zur Zuverlässigkeit des Modells fällen zu können. Die Weiterentwicklung neuer mobiler Geräte, wie Smartwatches oder Augmented Reality Brillen, bietet zudem weitere interessante Möglichkeiten für zukünftige Studien.

Abstract

Modern mobile devices offer a great variety of data that can be recorded. This broad range of informations offers the possibility to tailor applications more to the needs of a user. Several context informations can be collected, like e.g. information about position or movement. Besides integrated sensors, a broad range of additional sensors are available which can be connected to a mobile device. These additional sensors offer for example the possibility to measure physiological signals of a user.

The human body offers a broad range of different signals. These signals have been used in several examples to conclude on the state of a user. The different signals allow to get a deeper insight into emotional or mental state of a user. Electrodermal activity gives feedback about the current arousal level of a user. Heart rate and heart rate variability can give an estimation about valence and mental load of a user.

Several models exist to conclude from information like valence and arousal on different emotional states. Russell defined a two dimensional model, using valence and arousal to define affective states. Yerkes and Dodson developed a curve that expresses the relationship between arousal and performance of a user.

Different examples exist, that use physiological signals to determine the user state for tailoring and adapting of applications. At the time of this work most of these examples did not address the usage of physiological signals for user state estimation in mobile applications and in mobile scenarios. Mobile scenarios lead to several challenges that need to be addressed. Influencing factors on physiological signals, like e.g. movement have to be controlled. Furthermore a user might be interrupted and influenced by environmental aspects. The combination of physiological data and context information might improve the interpretation of user state in mobile scenarios.

In this work, we present a model that addresses the challenges of usage in mobile scenarios to offer an estimation of user state to mobile applications. To address a broad range of mobile applications, affective and cognitive state are provided as output. As input heart rate and electrodermal activity are used, as well as context information about movement and performance. Electrodermal activity is measured by a simple sensor that can be worn as a wristband. Heart rate is measured by a chest strap as used in sports.

The input channels are transformed to affective and cognitive state based on a fuzzy rule based approach. With help of fuzzy logic, uncertainty can be expressed and the data continuously being processed. At the start, input channels are fuzzified by defined functions. After a that, a first fuzzy rule set transforms the input signals into values for valence, arousal and mental load. In a second step, these values and context information are transformed with another fuzzy rule set to values for affective and cognitive state.

Affective state is based on the model of Russell, where valence and arousal are used to determine different emotional states. The output of the model are eight different affective states (alarmed, excited, happy, relaxed, tired, bored, sad and frustrated), which can have a high, medium, low or very low value as output. Cognitive state is determined based on mental load and context information about performance and movement. The output value can be very high, high, medium or low. The model was implemented as background service for Android devices. Different applications have been used for evaluation of the model. The model has been integrated in a multiplayer space shooter game, called "Zone of Impulse", which mainly benefits from the affective state. Cognitive state is more addressed in applications like a simple vocable trainer, which adapts difficulty based on user state.

A study to evaluate different aspects of the model has been conducted. The study was designed to investigate the suitability of the model for mobile scenarios. The game "zone of impulse" and the vocable trainer have been investigated in different configurations. Versions with integrated model have been compared to version of the applications without model, as well as versions of the model without context information.

In total 41 participants took part in the study. A part of the participants had to do the tasks of the study in a mobile scenario, walking around several streets. The remaining participants had to do the tasks in a controlled environment in a sitting position. Different aspects were collected with ratings and questionnaires.

Overall, participants rated that they did not feel impaired by the sensors they had to wear. The results showed, that the combination of physiological data and context information had an advantage against versions without context information in part of the ratings. A comparison between versions with and without model showed, that the subjective mental load ratings were significantly better for the version with model. Subjective ratings for aspects like fun, overstrain and support were mixed.

When comparing the application versions in indoor and outdoor scenarios, no significant difference could be found, which leads to the assumption that there is no loss of interpretation quality in outdoor scenarios. The results also showed that the model seems to be robust enough to compensate the loss of an input channel, as there was no significant difference between application versions with full integrated model and versions with one channel lost.

With the model developed in this work, context information and physiological data were combined to improve user state estimation. Furthermore pitfalls of user state estimation in mobile scenarios are overcome with this combination. However, the model has only been evaluated with a limited amount of applications and situations that mobile scenarios offer.

Contents

1. Introduction

1.1. Motivation

In the last years, mobile devices pervaded more and more areas of everyday life. Mobile devices get smarter and more powerful. In the past, mobile devices had only a limited functionality and were not small and lightweight. The usage of integrated sensors makes mobile devices smarter and extends the application scenarios, from just using them as a telephone to a multimodal communication assistant, which has integrated functionalities like e.g. navigation or playing music.

Applications from many different areas found their way onto mobile devices. Tasks like searching for the next public transport connection or a nearby restaurant can be done mobile. New devices like smartwatches or google glass, offer new possibilities in daily life. Recent surveys showed that in representative total population with age 16 and above 65% of people living in Germany use at least a smartphone [Cor15a].

In different applications, a broad range of information is used to tailor the application to the situation of the user, aiming for a better Human Computer Interaction (HCI). Assistants on mobile devices help to organize information to the needs of a user, based on different context information like location or entries in the calendar. Examples for existing systems are google Now[Inc15b], Microsoft Cortana [Cor15b] and Apple's Siri [Inc15a].

Besides context information, physiological signals allow to add new aspects to applications. Many researchers developed and evaluated concepts of using physiological signals for an interpretation of affective states, emotions and cognition. Nearly all examples of current research in this area, were developed and evaluated for controlled environments. When stepping out of the lab, additional aspects have to be considered.

Combining context information and physiological signals offers great potential. Applications would be able to not solely decide on the context, but also include the state of the users themselves for adaptation and assistance in applications. In this work these topics will be examined, concepts will be developed and finally evaluated with different applications.

1.2. Definitions and Limitations

This thesis covers a broad range of different topics and areas. The focus of this work is on the Human Computer Interaction area, some topics will be limited to the aspects addressed by HCI. Examples are emotions, affect and cognitive state. This work focuses on different defined categories of user state for HCI, based on psychological concepts of affect and emotion, which are used within recent state of the art research in the area of HCI. A detailed definition and overview of emotions as used in psychological research will not be given.

Context information is used in this work. Depending on the definition of context information in the literature, physiological signals are part of it or not. When talking about context information in this work, physiological signals are not included and will be looked at separately.

1.3. Thesis

Physiological signals can give an insight in the emotions, cognition and other parts of the current state of the user. In this thesis, physiological signals are examined and used as an input signal for mobile applications. Mobile scenarios vary, as the user can sit at home using a mobile device or be outside trying to catch a bus. Context information help narrowing down the current situation. As physiological signals are partially influenced by different aspects, like for example movement. These aspects have to be considered in modeling of user state.

The thesis will examine, how physiological signals can be used to conclude on the current state of a user for mobile applications. The questions, how different types of mobile applications can be supported and how interpretation in mobile scenarios can be handled, will be examined. To reach this goal different concepts and models used in recent research are examined and finally adapted to fit to the challenges of mobile scenarios.

In a second step, context information is examined and integrated to support and improve the developed model. The question if and how context information coming from mobile applications and mobile phone can improve interpretation quality, will be examined in this work.

Finally, the concept will be examined under aspects of robustness, as it may happen in mobile scenarios, that one or more input channels get lost due to bad connections or empty batteries. The question if a channel loss can be compensated without a big drop in interpretation quality will be investigated.

This work will introduce background information on used concepts and methods, needed for understanding the scope of this thesis in chapter 2. Chapter 3 will give an overview and an analysis of current research and state of the art concepts and applications. The results of this analysis is followed by the presentation and outline of the three theses, which are the foundation of this work in chapter 4. In the following chapters, an overview of the concept (chapter 5), the details of the model (chapter 6) and finally the implementation of the concepts (chapter 7) will be presented. Chapter 8 introduces several applications, which use the developed concepts. Different studies, supporting the thesis, are presented in chapter 9. Finally, the thesis closes with a conclusion and gives an outlook on future work in chapter 10.

2. Background

This thesis addresses several different research areas. For a better understanding of the thesis, an overview and introduction of the most important topics is given. This chapter starts with a brief introduction of physiological signals, the mechanisms of the nervous system followed by a detailed introduction of electrodermal and cardiovascular activity.

Several definitions, that will be needed for the interpretation of the used signals, like arousal, valence and mental load, are introduced accompanied by models and concepts for interpretation of user state. In the following subchapter, an introduction to context information and its definitions is given. Finally, methods and concepts for implementation of classification models used in current research are introduced.

2.1. Biology of Physiological Signals

To understand what physiological measures are and how they can be used, knowledge of biology of the human body is needed. The human body contains many signals that can be measured and quantified. Body functions are regulated by the nervous and the endocrine system. The endocrine system regulates functionalities like reproduction and digestion. It uses hormones for communication of signals via circulatory systems to their target. In comparison to the nervous system, it reacts slower. The nervous system on the other hand regulates functionalities, which need in most cases a quick reaction like receiving and responding to a stimulus from the environment. The nervous system regulates these functionalities by electrical signals and the release of neurotransmitter. [She08]

This chapter will take a closer look at the nervous system and its components. The two main divisions of human nervous system controlling physiological signals are the central nervous system (CNS) and the peripheral nervous system (PNS). The CNS consists of brain and spinal cord. PNS on the other hand, consists of different neurons, mainly sensory and motor neurons. Both systems are closely interconnected with each other as well as with the endocrine system. [She08]

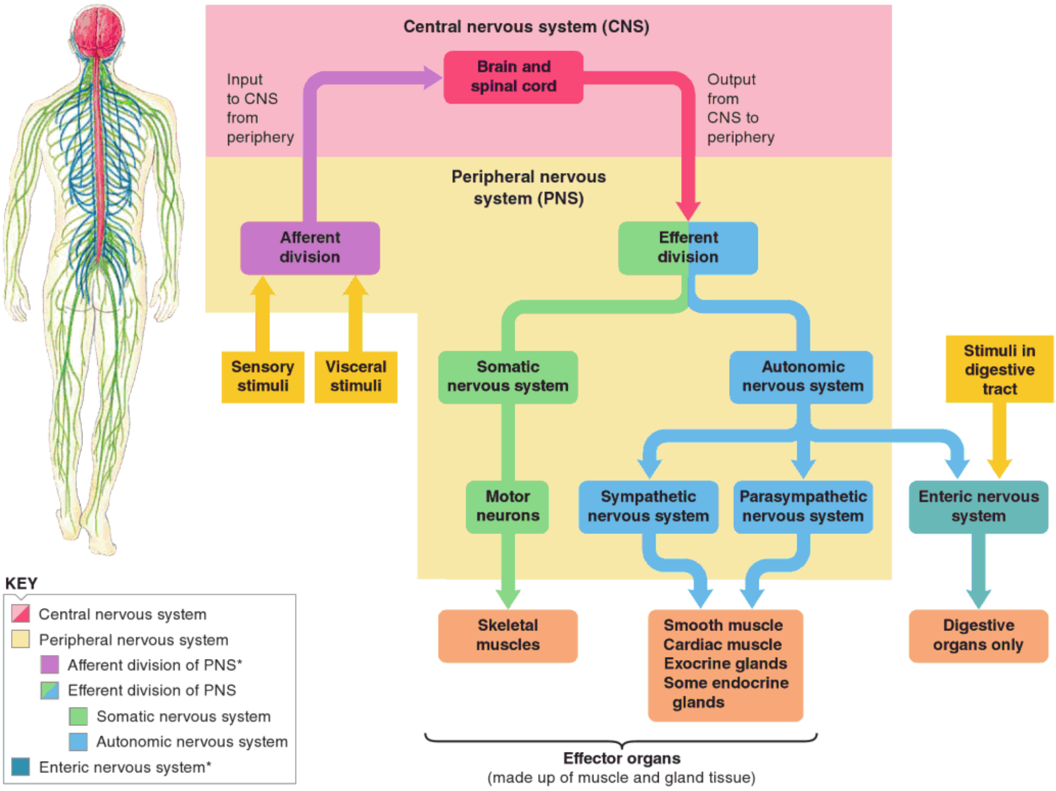


Figure 2.1.: Overview of the human nervous system and it several parts and functionalities in the human body [She08]

Figure 2.1 shows the subdivision and relationship of the different parts of the nervous system and the corresponding locations in the human body. As shown in the figure, the PNS can be further distinguished in different parts, which will be introduced in the following after a short description of CNS.

2.1.1. Central Nervous System

The CNS consists of the brain and spinal cord of the human body. The parts of CNS are shown in red in the human body on the left side of figure 2.1. In contrast to the PNS, the CNS is protected by the blood-brain barrier and bones of the head and spine. Its main tasks are the processing of incoming information from the PNS and transmitting orders to the efferent division of the PNS.

Measures of the CNS are for example electroencephalography (EEG), measures of brain metabolism (Positron emission tomography (PET)) and event-related potentials (ERP) [RSI98].

2.1.2. Peripheral Nervous System

The PNS consists of nerve fibers. It can be divided in an afferent and efferent part. The afferent part transmits signals and information of sensory stimuli of the environment to the CNS. Vice versa, the efferent part receives information, orders and signals from the CNS. [CR02]

The PNS is subdivided into the autonomic nervous system (ANS) and somatic nervous system (SNS). The SNS (shown in green in figure 2.1) covers functionalities that are responses to stimuli of the environment, e.g. regulate motor neurons to control movement of muscles. Most parts of the SNS can be controlled intentional by the human. The autonomic nervous system (shown in blue in figure 2.1) covers mainly functions of inner organs, like e.g. breathing and can mostly not be controlled intentional by the human. It is divided into sympathetic, parasympathetic and enteric nervous system. The enteric nervous system is mainly responsible for the digestive system. Sympathetic and parasympathetic nervous system are antagonists respectively to each other. Sympathetic nervous system goes hand in hand with a high alertness, attention and energy production. The impact on the human body are e.g. higher heart rate and inhibited digestion. On the other hand, parasympathetic system is connected with a relaxation and calm down of the body. Functionalities like heart rate and energy production slow down. [CR02]

Measures of the PNS are for example cardiovascular activity (autonomous part of the PNS), measures of the eccrine system (e.g. electrodermal activity for the somatic part of the PNS) and respiratory measures (autonomous part of PNS) [RSI98].

2.2. Electrodermal Activity

Electrodermal activity (EDA), often also referred to as Galvanic Skin Response (GSR) is the electrodermal reaction of the skin. The term electrodermal activity covers the electrical characteristics of the skin. EDA is involved in studies of many different research areas. Research about EDA began in the early 1900s by Vigouroux, who measured tonic skin resistance in 1879 and 1888 and Hermann and Luchsinger who examined innervation of cat sweat glands in 1878 [Bou92].

In the following physiology and the different components of EDA will be described as well as measurement and interpretation of EDA.

2.2.1. Physiology

The skin serves as a barrier between body and environment. It consists of different components and layers as shown in figure 2.2, for example sweat glands. Two different types of sweat glands exist, eccrine and apocrine sweat glands. Both types have different functions.

Eccrine sweat glands are primarily responsible for the regulation of body temperature. On the palm and plantar position, the eccrine sweat glands respond to psychological stimuli [Ede72], which is based on the high density of sweat glands on the hand [SMFC87]. The apocrine sweat glands are limited to different areas of the body and are less studied than the eccrine sweat glands. Their primary function is as well the regulation of body temperature. In contrast to the eccrine sweat gland, the apocrine sweat glands are not directly open on the surface of the skin [SMFC87].

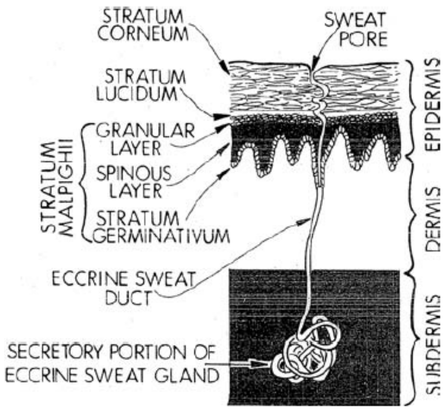


Figure 2.2.: Eccrine sweat gland [CTB07]

Figure 2.2 shows a profile of a eccrine sweat gland in the skin. The secretory part of the sweat gland lays in the subcutis. The duct connects the secretory part with the epidermis. Sweat, produced in the secretory part rises up the duct. When sweat fills the duct, the skin gets more conductive and the resistance of skin is lowered. EDA reacts within a time frame of 1 to 3 seconds after a stimulus appeared [CTB07]. Several studies showed that the sweat glands are connected to the sympathetic nervous system [CTB07].

2.2.2. Measurement

EDA can be either measured endosomatic or exosomatic. For exosomatic measurement a small direct or alternating current is used to measure conductance of skin. Endosomatic methods measure skin conductance without current. The output of the different measurement methods differentiate. Exosomatic measurement with direct current leads to skin resistance or conductance. Measured with alternating current, exosomatic measurement leads to skin impedance and skin admittance. Endosomatic measurement on the other hand leads to skin potential. In many studies, exosomatic measurement with direct current is the preferred method of measurement. [CTB07]

Several devices for measurement of EDA exist. Most devices have two electrodes which are placed on the palm of the hand. Electrodes are mostly made of silver/silver chloride

(Ag/AgCl) to minimize bias potential and polarization [CTB07]. Figure 2.3 shows possible placements for the electrodes. The most recommended electrode position is position 2. Another possible position for electrode placements are the feet.

Most studies use the non-dominant hand for measurement. This is motivated by the fact, that the skin of the non-dominant hand might show less skin lesion in comparison to the dominant hand. Furthermore, the dominant hand is free for other tasks in this case. [CTB07]

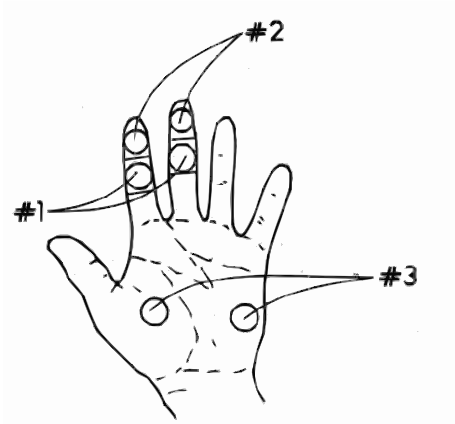


Figure 2.3.: Possible electrode positions for EDA measurement at the left hand [CTB07]

Measurement can be influenced by different aspects. One influencing factor is the actual condition of the skin. If a subject washes the skin with an abrasive soap, electrical properties of the skin might vary [VC73]. Therefore Venables [VC73] recommends to let subjects wash their hands before electrode placement with a non-abrasive soap. Besides condition of skin, measurement can also be influenced by humidity, ambient temperature and time of day. Several values of EDA can rise, with rising room temperature. [Bou92] recommends a room temperature of 23 Celsius and keeping humidity constant, if possible. Due to the issue that time of day can influence the values, these needs to be controlled in studies.

Different aspects of EDA can be measured, which are divided in phasic and tonic measures. The most used measures are Skin Conductance Level (SCL) and Skin Conductance Response (SCR). SCL is a tonic measure and reacts slower over time. SCR, on the other hand, counts to the phasic measures and reacts fast. Both will be described in Detail in the following.

2.2.3. Skin Conductance Level

SCL reacts within a time frame of 10 seconds to minutes [CTB07]. SCL is measured in microsiemens. The range is normally between 2 to 20 μS, when SCL is measured at the distal phalanges with exosomatic measurement and direct current.

When a new situation or stimulus is happening, SCL rises comparatively fast and decreases over time when at rest. Figure 2.4 shows the SCL of two different subjects. At the first 20 seconds, both subjects were at rest. After the rest period, three stimuli were presented. The curve shows the variation of SCL values between different subjects. Subject 1 starts at 10 µS, subject 2 at 5 µS. The curves also show the increase in SCL, when a stimulus is presented at 20, 35 and 50 seconds. The first time the stimulus was presented, the rise in SCL was bigger than the second and third time when the stimulus was repeated.

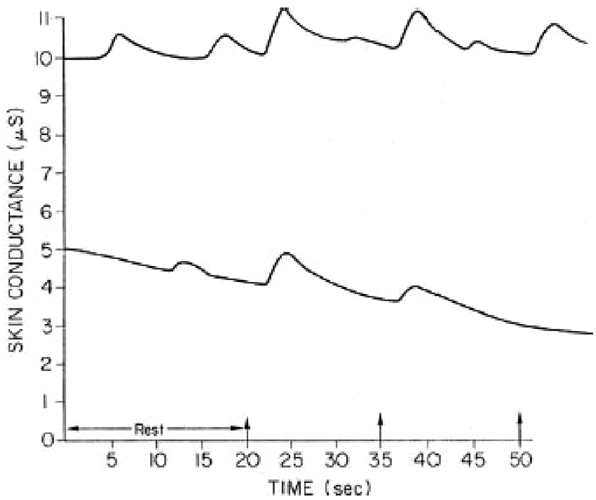


Figure 2.4.: Progress of Skin Conductance Level of two different persons [CTB07]

The measured values of a person cannot be compared with another person's values. A value of 1 µS might be a high value for one person, for another person it might be the minimum. Due to the individual differences, electrodermal activity has to be normalized. In this work, SCL is normalized by calculating the percentage of the overall SCL span:

$$SCL_{normalized} = \frac{SCL(t) - SCL_{min}}{SCL_{max} - SCL_{min}} * 100[in\%] \tag{2.1}$$

Minimum SCL values can be determined in a baseline measurement during a resting period. Maximum value can be determined over time or as proposed by [CTB07] initially by blowing up a balloon until it bursts. Interpretation might get more accurate with growing data set.

2.2.4. Skin Conductance Response

SCRs are elevations in form of small waves in the SCL. They are the phasic components of EDA. Figure 2.5 shows the course of a SCR. They can occur after a stimulus or spontaneously without a stimulus. When occurring without a stimulus, the SCRs are called Non-specific SCRs (short NS-SCR). If a SCR is a response to a stimulus, it occurs after a 1-4 seconds latency window after the stimulus occurred. [CTB07]

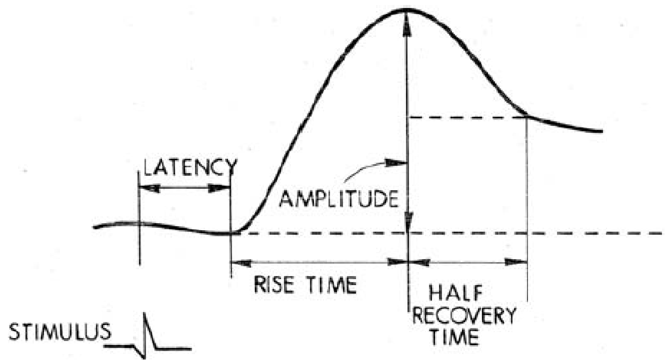


Figure 2.5.: Wave of a SCR [CTB07]

Different aspects of SCRs can be measured and analyzed. Minimum values for SCRs amplitude lay between 0.01 and 0.05 µS, to be interpreted as a SCR [CTB07]. After [CTB07], values for amplitude of a SCR normally range between 0.1 and 1.0 µS and have a rise time of 1-3 seconds. Magnitude and amplitude are two measures, which are commonly used and calculated. Magnitude is the average number of all SCRs of every stimulus presentation, even when there was not a response to the stimulus. For amplitude, on the other hand, only SCRs over the signals minimum value are used.

2.3. Cardiovascular System

The main part of the cardiovascular system is the heart, which is a muscle that regulates blood flow in the human body through several blood vessels. The circulatory system circulates blood through the body to transport oxygen, carbon dioxide, nutrients and blood cells to cells in the body. [CR02]

Several measures can be derived from the cardiovascular system, like for example blood pressure or heart rate. These measurements, their recording and interpretation are discussed in the following.

2.3.1. Physiology

The cardiovascular system is controlled by the sympathetic and parasympathetic systems of the autonomic part of the PNS. The heart itself consists of different chambers (shown in figure 2.6), which are electrically connected. During a diastole, the heart is filling up with blood, which gets pumped through the blood vessels in the systole.

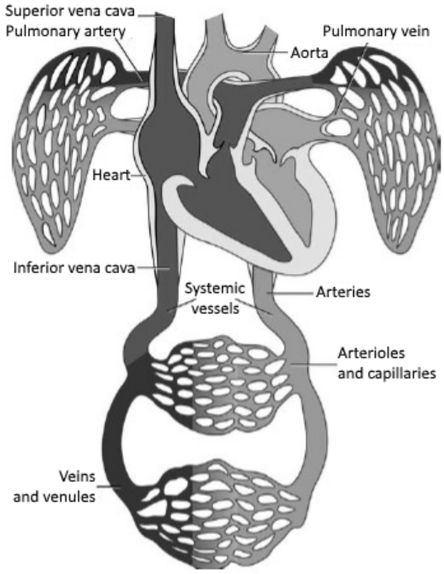


Figure 2.6.: Electrical activity of the heart [CTB07]

The electrical process of a heart beat can be recorded with an electrocardiogram (ECG) and consists of different characteristic (see figure 2.7): a P-wave, the QRS complex and a T-wave. During the end of a diastole, the P-wave is produced by the depolarization of the atrial node. When the atrial node contracts, a QRS complex follows on the P-wave. During QRS complex, ventral node contracts and at the end repolarizes, which can be seen as the T-wave. [CR02]

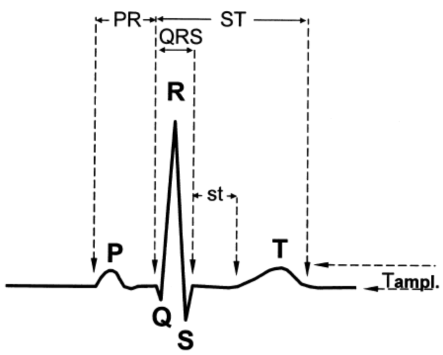


Figure 2.7.: QRS complex of a heart beat [CTB07]

2.3.2. Measurement

Activity of the heart can be measured by an electrocardiogram (ECG) in detail. It supports diagnosis of diseases and functional disorders of the heart. The in the previous subchapter introduced electrical processes can be measured by several electrodes. The first ECG was developed and measured by Einthoven in 1895 [Ein95].

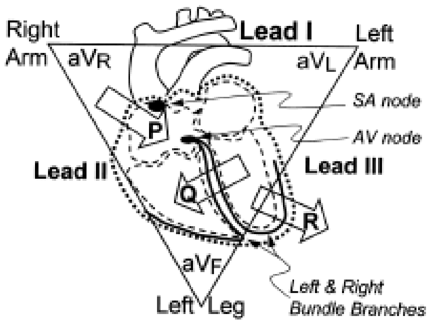


Figure 2.8.: Electrode placement after Einthoven [CTB07]

Several systems exist to place electrodes on a subject, depending on which aspects of the electrical process is desired to be measured. Figure 2.8 shows the electrode placement of Einthoven, placing the three electrodes at the right arm, the left arm and the left leg. With this bipolar lead, the potential difference between the arms and the leg are measured. The electrode placement based on Wilson is on the chest [CTB07].

In comparison to a full ECG, mobile heart rate monitors offer only limited information. Wireless heart rate monitors are mainly used in sports and measure heart beats per minute. First wireless heart rate monitor was published in 1983 by the company Polar [TBBM86]. Most heart rate monitors use a chest strap with two integrated electrodes, which measure the R-Peaks of the heart beats. Based on the R-Peaks, heart rate and RR-intervals can be calculated.

Other cardiovascular activity is measured with different devices, which do not base on the electrical activity of the heart. For example blood pressure is measured with an inflated arm cuff and blood or pulse volume with a plethysmograph.

2.3.3. Blood Pressure

Blood pressure describes the amount of pressure that is needed to push the blood through the circulatory system. Blood pressure in arteries is higher than in the veins. It is measured in millimeters of mercury (mm Hg) and can be measured systolic and diastolic. Systolic blood pressure is higher than diastolic, as systolic blood pressure is measured, when the heart contracts at a ventricle systole. Diastolic blood pressure is measured when the blood vessels return to their origin size. [CR02] A average value for a 20 year old adult at rest is around 120 for systolic and 70 for diastolic blood pressure [CR02].

Blood pressure can be influenced by different factors like age, weight and stress [SRQ01]. Continuous measurement of blood pressure data in real time is at the moment not possible, due to the in- and deflating of the arm cuff [Man08].

2.3.4. Heart Rate

The heart beats between 60 to 70 times a minute in an adult human during light activity. Within a minute, between 5 and 7 liters of blood are pumped through the circular system by the heart. [CR02]

Besides age, several other factors influence heart rate. Trained persons have usually a lower heart rate, than persons doing no exercise. The maximum heart rate is age-dependent and declines with increasing age. Heart rate is very variable between different persons. A stress test can determine the exact maximum heart rate. Several formulas exist, to calculate an average maximum heart rate based on age. The most common and widely distributed formula to calculate maximum heart rate was developed by Haskell and Fox [Kol01] in the early 1970s:

$$HR_{maximum} = 220 - Age \tag{2.2}$$

Tanaka, Monahan & Seals developed in 2001 a formula based on more than 18000 test subjects [TMS01]:

$$HR_{maximum} = 208 - (Age * 0,7) \tag{2.3}$$

Even though the formula of Haskell and Fox is more commonly used and widely distributed, the formula of Tanaka, Monahan and Seals reached better results.

For resting heart rate, the US National Health Institute published values for adults. Athletes have a resting heart rate between 40 and 60, other adults resting heart rate varies between 60 and 100 [oH13].

2.3.5. Heart Rate Variability

Heart Rate Variability (HRV) describes the variation of intervals between two heart beats. The difference between two heart beats is measured in milliseconds. HRV can be used to quantify the mental effort of a person [RSI98].

HRV can be analyzed by time-domain or with spectral methods. Malik et al. [MBC$^+$96] described several standards for the procedure of analysis. To calculate the power spectrum density, parametric and nonparametric methods can be used. Both offer different advantages and disadvantages. Auto-regression or a Fourier Transformation can be used.

The high frequency band ranging from 0.15 to 0.4Hz reflects activity of the parasympathetic parts of the ANS, the low frequency band between 0.04 and 0.15 Hz reflects sympathetic parts of ANS. The frequency band around 0.1Hz is used to determine mental effort [VTM87]. Typically, parts of the low frequency band are used, ranging from 0.06 to 0.14 Hz.

Figure 2.9 shows two examples for the result of a spectral analysis. Figure 2.9 (a) shows a relaxed person. Figure 2.9 (b), on the other hand, has a flattened low frequency band, which indicates a higher mental load.

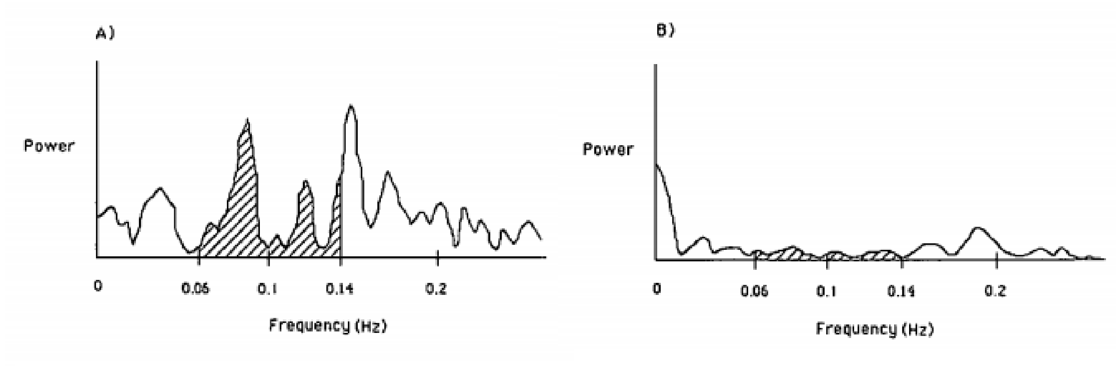


Figure 2.9.: Example for two HRV power spectra from [VTM87]. The left one shows a relaxed participant, the right one a participant under stress.

After spectral analysis Vicente [VTM87] proposes to integrate the power of the low frequency band (ranging from 0.06 Hz to 0.14 Hz) and normalize it with the average value from baseline measurement. By subtracting this result from 1, a value between 0 and 1 is the result, where 0 correlates to no and 1 to a high mental effort.

2.4. Mental Load

Mental load is a measurement, that is used in many areas. However, no unique definition of mental workload exists. Many definitions of workload include that mental workload is composed of the interaction between a task and different aspects of the user, like capabilities and motivation [Mor79] [Jex88] [Kra91]. Based on different definitions Kramer summarized workload in [Kra91] as:

> ″Mental workload has been defined as the ″costs″ a human operator incurs as tasks are performed.″

The ″costs″ described in the definition stand for the capacity a human operator is using for the task.

Workload does not necessarily rise with task difficulty. Tulga and Sheridan [TS80] reported that an increasing demanding task does not imply a higher level of workload. They conclude that user lowered their mental effort, as the task was getting too difficult to solve appropriately.

Despite the lack of an unique definition, mental workload is been used in many studies and applications. Mental load can be measured in different ways: subjective with rating scales, by performance measures or with help of physiological signals. In the following the different measurement methods will be described further.

2.4.1. Physiological Measures

Depending on the scenario and the aspects of workload, several physiological signals come into consideration for measurement of mental workload. Kramer [Kra91] examined different physiological signals as measurement for mental load under the aspects of sensitivity, diagnosticity, intrusiveness and reliability. Kramer stated, that different physiological signals measure different aspects of mental workload. Part of the study were event-related potentials measured from the brain, cardiovascular activity, pupil diameter and measures of respiration. EDA turned out to be only useful to identify shifts between situations of different kinds of workload. Event-related potentials achieved highest diagnosticity to determine mental workload. On the other hand, measurement of ERP is intrusive in comparison to other measures.

Changes in pupil diameter proved to be a reliable measure for mental workload. Kramer [Kra91] recommends to use pupil diameter measurement only in controlled experimental conditions, where the head movement can be controlled, due to the difficulty of fast and exact determination of diameter changes. Since the study of Kramer, measurement devices for pupil diameter improved dramatically, allowing a fast measurement with eye-tracking systems even in situations, where the user is moving the head, e.g. during driving a car [PKSH10]. However, pupil diameter measurement requires a user to have some sort of camera in direction of the eyes.

Cardiovascular measures in the study of Kramer [Kra91] were blood pressure, blood volume and ECG. As described in chapter 2.3.5 HRV can be used for assessment of mental workload. Meshkati [Mes88] examined besides HRV several other physiological signals regarding their usage for mental load measurement and calls it the most promising measurement for workload.

2.4.2. Subjective Ratings

The most famous rating scale for mental workload assessment is the NASA Task Load Index (NASA-TLX) originally designed for aviation which was introduced by Hart in

[HS88]. The NASA-TLX is a multidimensional scale to obtain workload of a person. The NASA-TLX is divided into six different subscales, which are shown and further described in table 2.1. Each subscale is divided into 20 items to allow a rating between low and high or good and poor. Before answering the six scales (mental demand, physical demand, temporal demand, effort, performance and frustration level), user have to fill out a weighting for the six aspects to rate which of the scales influences them more in comparison to the other scales. Based on the results of these weighing, the results of the six scales are calculated. In the end, all six scales are added together for the mental workload.

Title	Endpoints	Descriptions
Mental Demand	Low/High	How much mental and perceptual activity was required (e.g. thinking, deciding, calculating, remembering, looking, searching, etc.)? Was the task easy or demanding, simple or complex, exacting or forgiving?
Physical Demand	Low/High	How much physical activity was required (e.g., pushing, pulling, turning, controlling, activating, etc.)? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?
Temporal Demand	Low/High	How much time pressure did you feel due to the rate or pace at which the tasks or task elements occurred? Was the pace slow and leisurely or rapid and frantic?
Effort	Low/High	How hard did you have to work (mentally and physically) to accomplish your level of performance?
Performance	Good/Poor	How successful do you think you were in accomplishing the goals of the task set by the experimenter (or yourself)? How satisfied were you with your performance in accomplishing these goals?
Frustration Level	Low/High	How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?

Table 2.1.: Rating scale definitions of the NASA-TLX [Har06]

These six subscales address the challenge of a clear workload definition. The six aspects of workload were designed to meet most people's experience of mental workload [Har06].

In [Har06] the usage of NASA-TLX has been examined over the last 20 years. In total 550 studies from different countries were reviewed by Hart [Har06] regarding different aspects like study environment, focus of the study etc. The results showed, that visual and or auditory displays were focus of most studies with a portion of 31 %. Looking on the environments of the studies, Air Traffic Control (10), civilian (12) and military cockpits (5) had the biggest portions. Computer users had a portion of 7 and user of portable technologies like smartphones had a portion of 4.

Besides the NASA-TLX several other rating scales for mental workload exist, like for example the Subjective Workload Assessment Technique questionnaire (SWAT) [RN88]. The SWAT questionnaire consists of an additive multidimensional representation of three dimensions, like shown in figure 2.10. The dimensions are: time load, psychological stress and effort load.

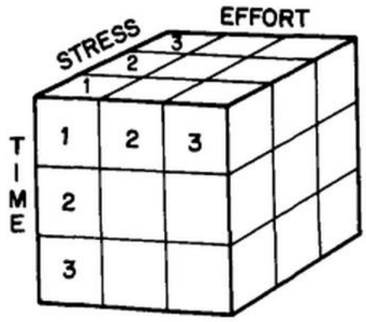


Figure 2.10.: Three-dimensional workload model of Reid and Nygren [RN88]

Each of these dimensions is divided into three level: low, medium and high. The single definitions of each level for each scale is shown in table 2.2. To assess workload with SWAT, two steps have to be conducted. The first step, called the scale development, is based on 27 cards, which contain all possible combinations of the three scales and its three levels. The participant is asked to sort the cards in order of increasing workload by own perception. In the next step, the participant rates its workload on the scales. Finally, each rating is converted to a value between 0 and 100, based on the scale of step 1.

Luximon et al. [LG01] developed a simplified version of the SWAT questionnaire to handle pitfalls like missing sensitivity of low workloads. They compared different types of simplification of the card sorting step. Card Sorting was compared with SWAT without card sorting and a method of pair wise comparison.

The results showed, that the approach of SWAT without card sorting was most sensitive, pairwise comparison moderate to more sensitive and the full card sorting process lead to

I.	**Time Load**
1.	Often have spare time. Interruptions or overlap among activities occur infrequently or not at all.
2.	Occasionally have spare time. Interruptions or overlap among activities occur infrequently.
3.	Almost never have spare time. Interruptions or overlap among activities are very frequent, or occur all the time.
II.	**Mental Effort Load**
1.	Very little conscious mental effort or concentration required. Actively is almost automatic, requiring little or no attention.
2.	Moderate conscious mental effort or concentration required. Complexity of activity is high due to uncertainty, unpredictability, or unfamiliarity. Considerable attention required.
3.	Extensive mental effort and concentration are necessary. Very complex activity requiring total attention.
III.	**Psychological Stress Load**
1.	Little confusion, risk, frustration, or anxiety exists and can be easily accommodated.
2.	Moderate stress due to confusion, frustration, or anxiety noticeably adds to workload. Significant compensation is required to maintain adequate performance.
3.	High to very intense stress due to confusion, frustration, or anxiety. High extreme determination and self-control required.

Table 2.2.: Rating scale definitions of SWAT [RN88]

least sensitivity.

Both introduced questionnaires, NASA-TLX and SWAT, have been widely used, especially in determination of workload in aircraft multitask situations [RDMP04]. However, subjective ratings require that a user fills out a questionnaire or answers question. During interaction with a system, this might lead to an interruption in workflow.

2.4.3. Performance Measures

Besides subjective ratings and physiological signals, mental load can also be estimated by performance measures like error rate, click rate or time span to solve a task. Task performance can be measured by a primary-secondary-task scenario [Lin91]. User focus on a primary task, while a secondary task is offered for situations of low workload in the primary task. Depending on how well users perform at the secondary task, workload can be estimated.

This approach is more objective as the subjective ratings, but offers only limited accuracy in the determination of different workload levels. Furthermore performance measure are not reliable in every situation, a person can be distracted from the environment or occupied with multitasking switching between different other tasks. If a person does not engage enough with the secondary task, workload estimation is very inaccurate or not possible.

2.5. Arousal

Arousal is defined as a state of high (excited, stimulated, awake) or low (calm, sleepy) activity. The terms tension-relaxation or activation are also commonly used in the sense of arousal. Arousal is widely used in different psychological concepts and models. It was first introduced in 1912 by Wundt [WP12], who proposed tension-relaxation as a dimension for describing emotions.

Many psychological models for the interpretation of emotions, affect and state of a person have arousal as one of their parts. Some of them will be described in chapter 2.7 in detail. In the following, methods of measuring arousal, physiological or subjective, will be introduced.

2.5.1. Physiological Measures

Different physiological signals can be used to conclude on the current arousal level. Changes in Skin Conductance Level correlate to arousal, as well as in parts of measured EEG [BL00]. A rising Skin Conductance Level, corresponds to a rising level of arousal. In the EEG the alpha waves (10-13 Hz) correlated to low arousal as well as an increase in frequency correlated to an increase in arousal [BL00].

Besides EDA and EEG, pupil diameter is a measurement for arousal [BMEL08]. Studies of Bradley et al. [BMEL08] showed, that pupil response correlated with emotional arousal and covaried to Skin Conductance Level. Decreased blood volume pulse and increased heart rate have also shown to correlate to arousal [MA07].

2.5.2. Subjective Measures

Different subjective measures exist, that have a scale for arousal in one or another way integrated. The self-assessment questionnaire manikin (SAM), developed by Lang et al. [BL94] is a questionnaire consisting of three single scales for arousal, dominance and valence. The items of each scale are presented as drawn pictures, offering a nonverbal possibility of assessing the different values. The scale for arousal is shown in figure 2.11. High arousal is represented by a figure implying movement and high alert. Low arousal on the other hand is represented by a sleeping figure.

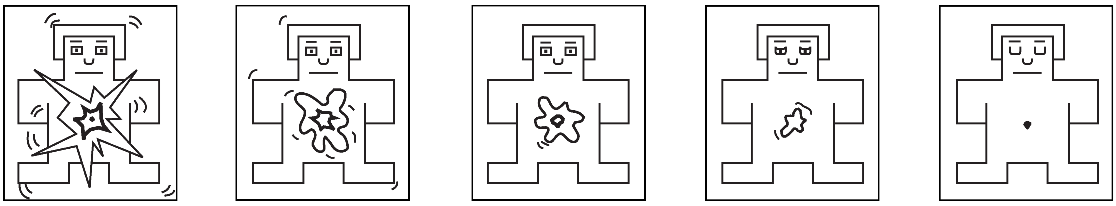


Figure 2.11.: Arousal scale of the SAM questionnaire [BL94]

Besides SAM, other questionnaires address arousal in different ways. For example the Affect Grid, which will be introduced in detail in subchapter 2.7.2, has no single scale for arousal as it is integrated in a two dimensional scale of valence and arousal.

2.6. Valence

Valence describes whether a situation or feeling is positive or negative. The term pleasure is also commonly used in the same way as valence. Frijda [Fri86] defines valence in the following way:

> "Events, objects, and situations may possess positive or negative valence; that is, they may possess intrinsic attractiveness or aversiveness"

Valence has been used in many concept and models to describe affect [RWM89]. In the following, different concepts of measuring valence with physiological signals or subjective measures are introduced.

2.6.1. physiological measures

Different physiological measures can be used, to conclude on the level of valence. One commonly used measure is the analysis of facial expression with help of an electromyogram (EMG). Electrodes for EMG are placed near the cheek and at the forehead to measure smiling and frowning [MA07].

Besides EMG, heart rate, irregularity of respiration and pupil diameter have shown potential to determine level of valence [MA07].

2.6.2. Subjective Measures

Similar to arousal, different subjective measures exist, that have a scale for valence but do not measure valence solely. The SAM questionnaire also offers a scale for valence, shown in figure 2.12. In this case, the figure ranges from a smiling to a depressed looking figure.

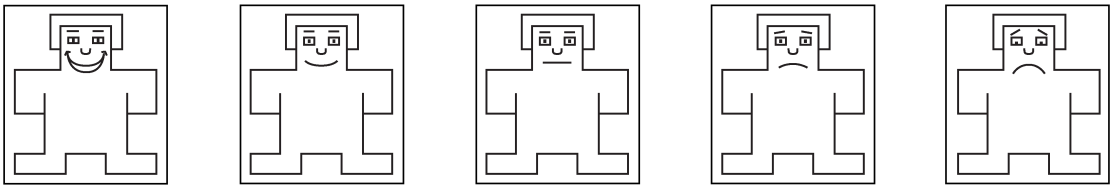


Figure 2.12.: Valence scale of the SAM questionnaire [BL94]

Besides the SAM questionnaire, the affective grid has also a dimension, which covers valence on a nine point scale for valence.

2.7. Modelling Psychological User State

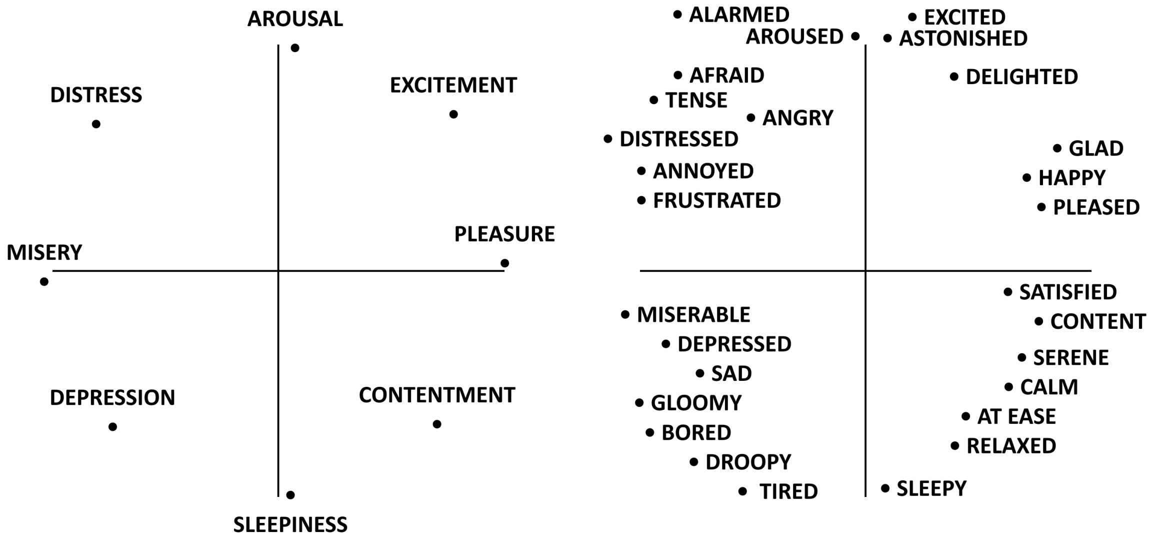
No clear definition of emotion and how to measure it exists [WR10] [KJK81]. Different models exist, to further describe the psychological state, mostly based on valence and arousal or similar concepts [YRS11]. The models describe different emotions or different aspects of an emotional or affective state. As this thesis concentrates on the aspects of computer science, only a short overview of the most commonly used models within Human Computer Interaction (HCI) is given.

Russell developed models to describe the affective state of a person (see 2.7.1 and 2.7.2) and the core affects (2.7.3). Yerkes and Dodson defined the Yerkes-Dodson law, which describes the process of performance of a person (see chapter 2.7.4). Csikszentmihalyi defined the flow zone, communicating the ideal zone between skill and challenge (see chapter 2.7.5). These models and concepts are further described in the following subchapters.

2.7.1. Russell's Circumplex Model of Affect

The circumplex model of affect based on Russell [Rus80] (figure 2.13(a)) is a two dimensional model, describing different states of affect by a linear combination of valence (x-axis) and arousal (y-axis). The arousal is a description of the activation level of a person. Valence describes if the feeling is pleasant or unpleasant. Due to the valence and arousal axis, the model is also often called valence-arousal space.

The first version of the model, with 8 different emotions is shown in figure 2.13(a) as it was predicted by Russell before testing, with the affects pleasure (0°), excitement (45°), arousal (90°), distress (135°), displeasure (180°), depression (225°), sleepiness (270°) and relaxation (315°) on the circle.



(a) Predicted model of Russell [Rus80]

(b) Circumplex model with 28 words [Rus80]

Figure 2.13.: Left: first version of circumplex model [Rus80], right: model based on several studies [Rus80]

Based on several studies, Russell determined the position of 28 different emotions in the circumplex model. The results showed, that the original predicted model with the eight different affects corresponded to the most of the different scaling variations of the 28-word version of the model (shown in figure 2.13 (b)).

The original circumplex model, with 8 different affective states is especially suitable for applications like games. The extended model with 28 different affective states on the circumplex model might be to fine granulated.

2.7.2. Russell's Affect Grid

The affect grid itself is not a model for emotion or affect. The affect grid developed by Russell [RWM89] is designed to offer a one-item scale for determination of current affect state. The affect grid is a 9x9 grid with the axis unpleasant-pleasant and high arousal - sleepiness shown in figure 2.14. Pleasant-unpleasant and high arousal - sleepiness correspond to the scales, which are often used for two dimensional models to describe emotions or affect, like the two models of Russell described in the previous subchapters 2.7.1 and 2.7.3.

Studies of Russell et al. [RWM89] showed, that the affect grid had validity and reliability in studies, where current mood or affect were recorded with help of the grid.

The grid was designed to easily access the state of a person, for a broad range of applications and studies. Application areas of the affective grid are manifold. Examples are

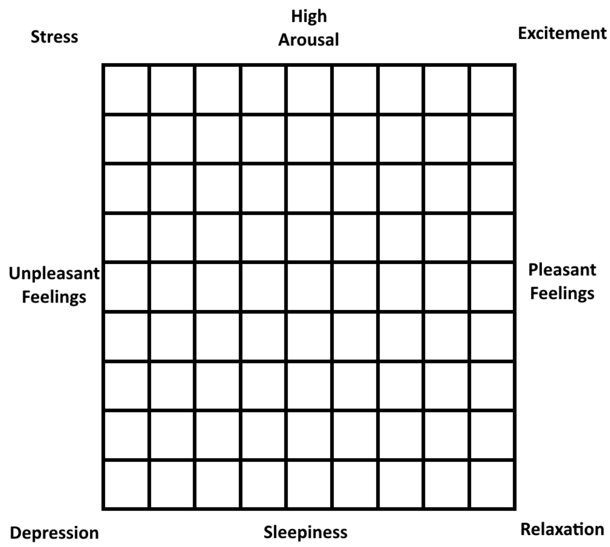


Figure 2.14.: Affect grid by Russell [RWM89]

studies about music [HTS02] or games [MA07]. As the affective grid helps to determine the affective state, it is a measure, that has its strengths in applications, which aim for a certain affective state like games, instead of controlling performance level.

2.7.3. 12-point affect circumplex

The 12-point affect circumplex (12-PAC) was developed by Yik et al. [YRS11]. The model is based on 12 different core affects. Yik et al. [YRS11] define core affect as "the simplest feeling" which cannot be further reduced to anything simpler in psychological terms. Core affects are a part of a mood and a person always has one of the core affects.

Figure 2.15 shows the 12 different affects, a combination of activation and pleasure level, with their position and angle at the circle. Each 30° one core affect is denoted. The figure also shows for each core affect a possible emotional state, e.g. sad or gloomy for a state of deactivated displeasure. The angles are only an estimation, as they cannot be determined exactly. Studies presented by Yik et al. [YRS11] show that the model was robust and is highly correlated with other mood scales. [YRS11]

As the model has basic definitions of the different affective states in comparison to the circumplex model, it is suitable to express an emotion or affect by its basic components.

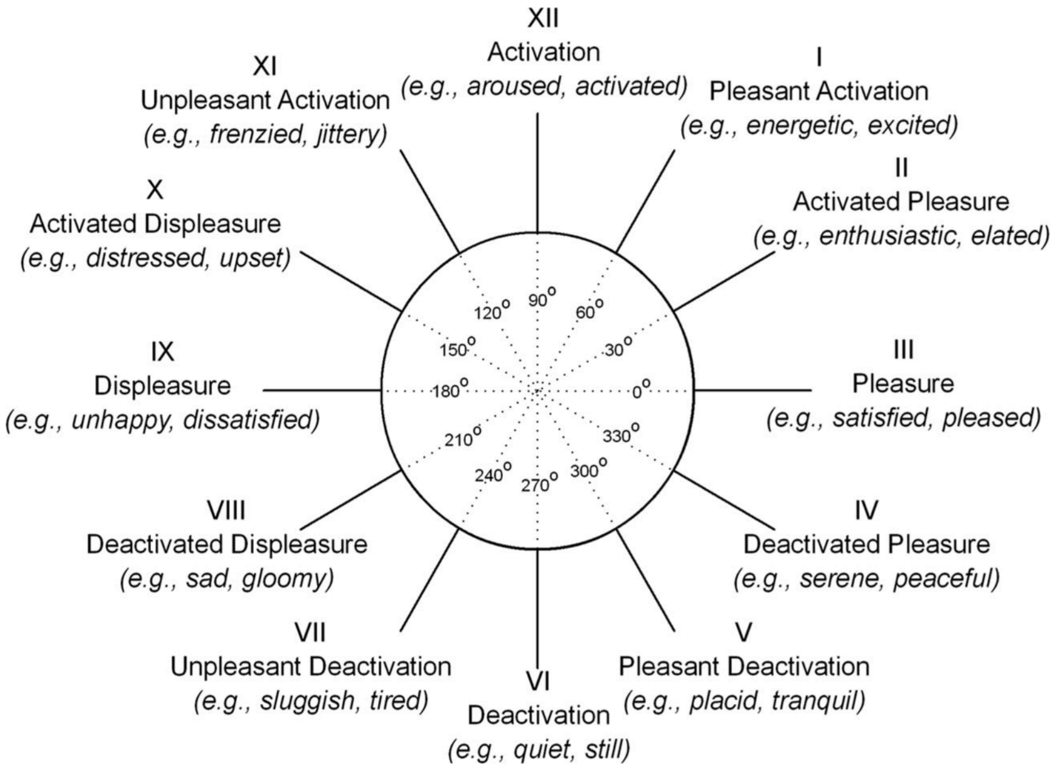
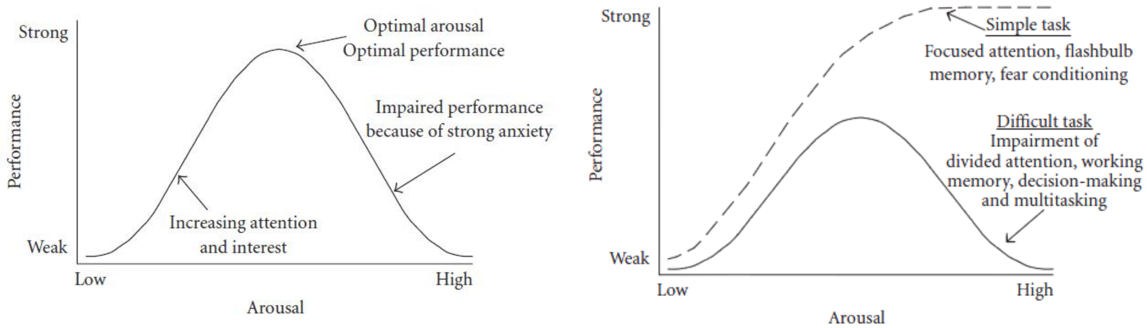


Figure 2.15.: Circumplex model divided into 12 core emotions [YRS11]

2.7.4. Yerkes-Dodson Law

In 1908 Yerkes and Dodson defined the Yerkes-Dodson law, based on results of their studies [YD08]. The law describes the relationship between performance and arousal in a curve shown in figure 2.16(b).

The curve shown in figure 2.16(b) distinguished between two different situations: simple and difficult tasks. During a simple task, performance stays high when reaching a certain level of arousal. In this scenario, subjects are able to focus the attention on the task. On the other hand, during a difficult task, subjects only achieve high performance during a mid level of arousal. Higher arousal values lead to a decrease in performance. In this situation, the resources of a person are overextended. Typical examples for this situation are multitasking and divided attention.



(a) Curve of Hebb (b) Yerkes-Dodson curve

Figure 2.16.: Left: Hebbian version of Yerkes-Dodson curve [DCP+07], right: Yerkes-Dodson curve based on current studies [DCP+07]

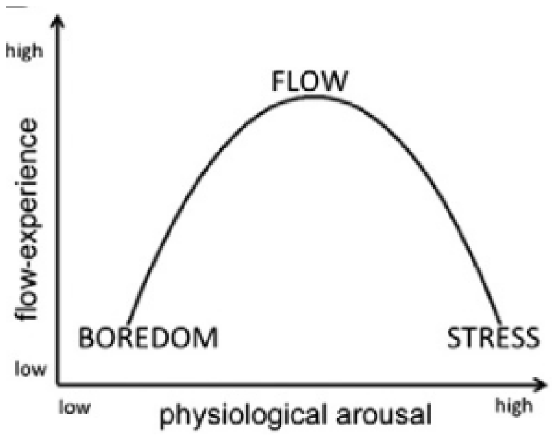
Figure 2.16(a) shows a result of the work from Hebb [Heb55]. Based on different studies, he came to the same results as Yerkes and Dodson. The difference between both researches is the differentiation based on task difficulty. Hebb did not distinguish between simple and difficult tasks to describe the relationship of arousal and performance.

Applications areas for the Yerkes-Dodson law are especially learning and tasks, where a certain performance level should be kept. The difficult task of the original Yerkes-Dodson law compares to multitask situations in mobile scenarios, where user play a game or use an application, but also have to pay attention to the environment.

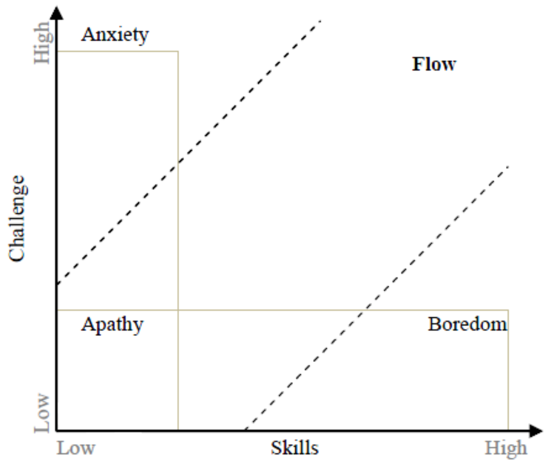
2.7.5. Flow

Csikszentmihalyi defined the original Flow model in 1991, which describes the relationship between challenge and skill [Csi91][NC02]. The relationship is visualized in figure 2.17 (b). The flow zone represents the optimal state between challenge (y-axis) and skill (x-axis) requirements. If the challenge is to low, the user is in a state of boredom. On the other side, if the challenge is to high, the user experiences a state of anxiety. Csikszentmihalyi defines flow as a state in which a person enjoys and is totally involved in the activity [Csi91].

Figure 2.17 (b) shows the already revised model by Ellis et al. [EVM94]. Ellis et al. revised the model by adding an apathy zone, when challenge and skills are low. This revised model is especially suitable for games. Typical application areas for this model are for example situations, in which the action itself is the goal instead of achieving a given performance level. Examples are games, drawing and sports. [NL08]



(a) Revised model by Ellis et al. from [NL08]



(b) relation between arousal and flow [PSS+14]

Figure 2.17.: Models for the relation between arousal and flow

Flow can be measured in different ways. A questionnaire, the flow state scale developed by Jackson and Marsh [JM96], quantifies the flow experience of a person in a subjective way. Physiological measures allow a objective estimation of flow experience. Peifer et al. [PSS+14] described the relationship between flow experience and arousal in a upturn U curve 2.17 (a). With low arousal user experience boredom, under high arousal they experience stress. In a mid level of arousal, flow experience achieves high values.

2.8. Context

Several definitions of context exist. One of the most used definition of context is from Abowd et al. [AD99], who define context as:

> ”Context is any information that can be used to characterize the situation of an entity. An entity is a person, place, or object that is considered relevant to the interaction between a user and an application, including the user and application themselves.”

Based on this definition, every information that leads to a closer description of the situation of a user is defined as context. In the following context sensitivity and context information and its categorization will be described.

2.8.1. Context Information

Context information can be distinguished between explicit and implicit gathered context information. [AD99] gives as an example for implicit context the identification of a user via vision based approaches, where on the other hand identification by a login form would be explicit information.

Type of context	*Available sensors*
Light	Photodiodes, colour sensors, IR and UV-sensors etc.
Visual context	Various cameras
Audio	Microphones
Motion, acceleration	Mercury switches, angular sensors, accelerometers, motion detectors, magnetic fields
Location	Outdoor: Global Positioning System (GPS), Global System for Mobile Communications (GSM); Indoor: Active Badge system, etc.
Touch	Touch sensors implemented in mobile devices
Temperature	Thermometers
Physical attributes	Biosensors to measure skin resistance, blood pressure

Figure 2.18.: Table of physical sensor types, that are widely used in mobile devices [BDR07]

Indulska and Sutton [IS03] distinguish between three different sources of data, to collect context information: physical, logical and virtual sensors. Physical sensors collect physical data. Figure 2.18 shows a broad range of physical sensors that are widely distributed. Virtual sensors on the other hand collect context data from applications or services, e.g. mouse-clicking rate. The so called logical sensors use the data collected by physical and virtual sensors, as well as other data sources to conclude on certain information. [IS03]

Different categories can be defined for the type of context information. The most commonly used are the categories defined by Ryan et al. [RPM98] and Day et al. [AD99]. Ryan et al. [RPM98] define the categories location, environment, identity and time. Dey et al. [AD99] also define the categories location, identity and time, but changed environment to the category activity. They motivate that with the fact, that activity gives an information about what is happening, where on the other side they argue that environment is only another description of context itself.

Context information can further be distinguished in primary and secondary context information [AD99]. Primary context information are for example information about the identity of a person. Secondary context information can be collected based on primary context information, e.g. contact details of a person based on the identity. [AD99]

2.8.2. Context Awareness or Sensitivity

Context awareness or sensitivity is defined as the capacity to adapt to a situation or environment based on information about the context. A first definition was given in 1994 by Schilit et al. [ST94]:

> ”Context-aware computing is the ability of a mobile user’s application to discover and react to changes in the environment they are situated in.”

Dey [Dey01] defined a context-aware system in 2001 more general comparing to [ST94], to fit to a broader range of applications :

> ”A system is context-aware if it uses context to provide relevant information and or services to the user, where relevancy depends on the user’s task.”

In addition to this definition Dey [Dey01] defined three different categories of context-aware application types, that were also kept as general as possible. The three categories are: presentation of information, automatic execution and tagging of context. Presentation of information offers for example the possibility to present information or services to a user, tailored to the current context. Automatic execution covers applications that might cover some kind of automation depending on the current context, e.g. the current location. Tagging of context with information can be used for similar future situations. [Dey01]

2.9. Concepts of User State Modelling

Different concepts exist in computer science for classification or estimation of an user state. In this subchapter, the background and foundations for the most commonly concepts used in the modeling of user state based on physiological data will be introduced. At first, a short overview of neural networks will be given, followed by an introduction of support vector machines, Fuzzy logic and Bayesian networks.

2.9.1. Neural Networks

Neural networks are based on neurons. The neurons of a neural network are inspired by the neurons of the human brain, which collect, process and transmit signals. McCulloch and Pitts [MP43] described a neuron mathematically. An example for a neuron for neural networks is shown in figure 2.19:

Figure 2.19 shows the input and output links, as well as the parts of the neuron itself, the input and activation functions and the output value. Every input link has a weight W. These input weights get summed up by the input function. The activation function

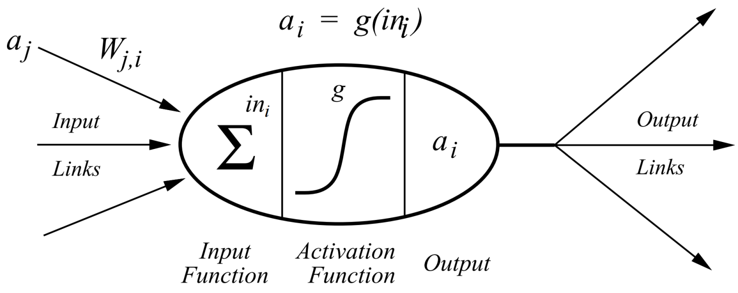


Figure 2.19.: Russell-Norvig Neuron [RN04]

is applied to the result of the input function to determine the output a_i. The activation function can be a threshold or a sigmoid function. [RN04]

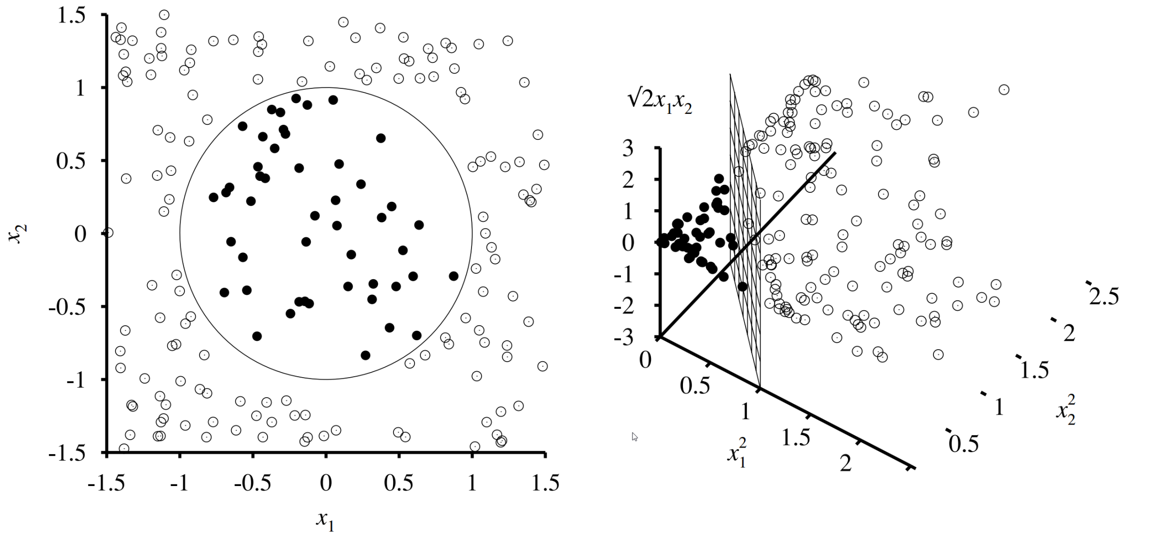
The topology of the networks can be cyclic (Feedforward networks) or acyclic (Recurrent networks). Cyclic networks feed their output back to the input. Neural networks can have single or multiple layer. In single layer networks, input neurons are directly connected with the output neuron. Single layer networks can represent linear functions, multi-layer networks on the other hand are more expressive and can represent nonlinear functions as they have several layer of neurons between input and output neurons. [RN04]

Neural networks are able to learn. Different learning algorithms for neural networks exist. Most learning algorithms are based on the principle of back propagation with gradient descent for optimization. To train a neural network with backpropagation, training sets are needed, to compare the output of the training sets with the correct result. The error between correct result and output of the neural network are then compared and propagated back through the network to adjust the weights between the neurons. [RN04]

Single layer networks allow efficient and fast learning. The more layer the network has, the more complex gets the efficient solution of the learning process with backpropagation. [RN04]

2.9.2. Support Vector Machine

The support vector machine (SVM), or also called Kernel machines, is a fast and efficient method for classification. They offer the same expressiveness as neural networks, but have a more efficient learning algorithm. SVMs are based on the idea, that different objects of a set get divided by vectors into different classes. These vectors have support vectors, which run parallel to the dividing vectors through the closest point of each divided class. To determine support vectors, a training data set is needed where it is known which data belongs to which class. Figure 2.20 shows an example for determination of dividing vector and support vectors. [RN04]



(a) Division into two classes in two dimensional space (b) Transformation of solution into three dimensional space

Figure 2.20.: Transformation of classes for preparation of vector calculation [RN04]

Figure 2.20 (a) shows an example of two classes (black and white dots), that should be divided into two separate classes. The shown solution is a circle and not linear. To solve the problem, the function is transformed into the three dimensional space (x_1,x_2,...). The nonlinear border between the two classes of figure 2.20(a) is now a linear border. When transformed back into two dimensional space, by projection to the axis x_1 and x_2, a vector shown in figure 2.20, is the result.

On both sides of the vector, a parallel supporting vector is determined, going through the closest point of the set to the vector (see figure 2.21). This supporting vectors allow a fast and efficient determination in which class a point belongs to.

Depending of the number of dimensions that are needed to calculate the support vectors, the computation gets complex. To handle this problem, Kernel functions are used which have a high performance in calculation. Depending on the problem, different Kernel functions can be used. An overview of Kernel functions and details of their functionality is given in [STC04].

SVMs were examined in different use cases. Russell and Norvig [RN04] list recognition of hand written numbers as an example. In this example, SVMs were compared to approaches with other approaches like neural networks. The virtual SVM had a higher classification rate as Neural networks approaches in the study.

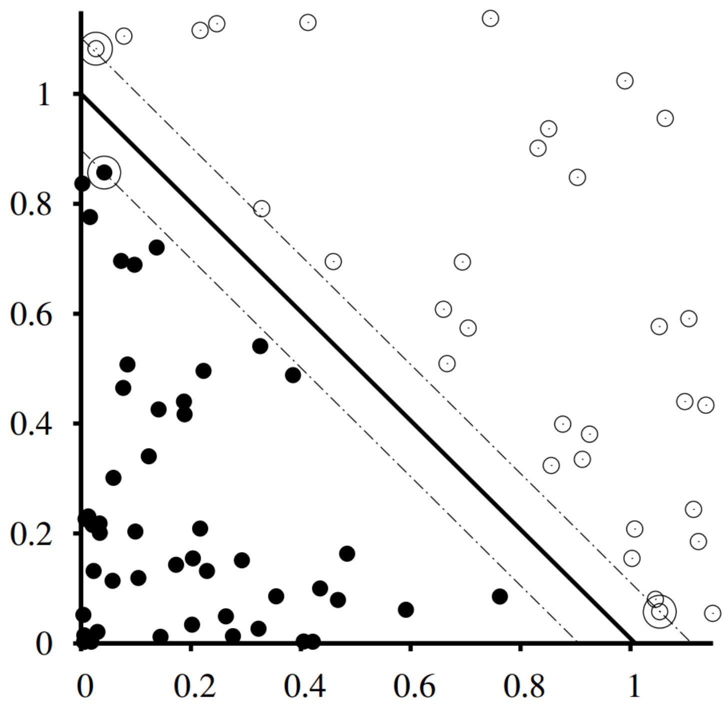


Figure 2.21.: Transformation back from three to two dimensional space followed by determination of supporting vectors (dashed lines) [RN04]

2.9.3. Fuzzy Logic

Zadeh [Zad65] introduced Fuzzy sets and fuzzy logic in 1965. Fuzzy logic is based on the idea to describe sets where an element can not only be in or outside of a set, but also be partially a member of a set. This allows the handling of imprecision for different applications.

A typical fuzzy system is based on three steps: fuzzification of input, logic processing and defuzzification of output. In the first step, the input is fuzzified and represented as a fuzzy set. Fuzzy sets are defined by membership functions, which assign each point of the set a value between 0 and 1. This value describes the grade of membership to a certain set. Figure 2.22 shows two examples for membership functions, showing temperature and pressure. Different fuzzy regions are defined, for example cold, cool, tepid, warm and hot for temperature. Depending on the input value on the x-axis, degree of membership can be determined (y-axis). E.g. if pressure is around 1200 a degree of membership for "OK" and "Strong" is returned, as the result is within both sets.

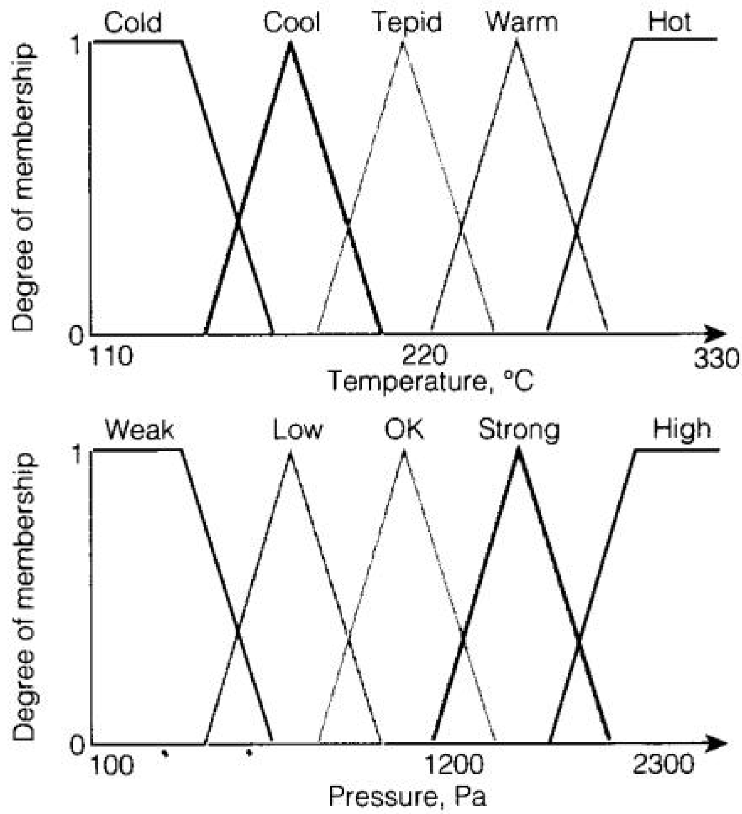


Figure 2.22.: Example for membership function for temperature and pressure [Cox92]

During logical processing, fuzzy rules are applied to transform the fuzzified input to the fuzzified output. Rules are based on IF and THEN. Additionally logical operators AND, OR and NOT are defined, called Zadeh operators. OR defines the maximum, AND the minimum and NOT the complement [RN04]:

$$AND : x \wedge y = min(x, y) \tag{2.4}$$

$$OR : x \vee y = max(x, y) \tag{2.5}$$

$$NOT : \neg x = 1 - x \tag{2.6}$$

An example for a rule is "IF temperature is cold THEN pressure is strong". The results of rules are the degree of membership to the certain fuzzy sets.

In a last step, the result of the logic processing is defuzzified. Different methods for defuzzification exist to choose a value out of the different possible solution sets and the degree of membership. During defuzzification, each variable is mapped to a corresponding crisp value out of it's solution fuzzy sets. For example, for the pressure value 1200 from figure 2.22 defuzzification decides if the result is ”Ok” or ”Strong”. Van Leekwijck and Kerre list several defuzzification methods in [VLK99].

In contrast to neuronal networks, the membership functions have to be created and can not be learned. Zadeh [Zad96] states, that the main purpose of fuzzy logic in comparison to other methods like Bayesian and neural networks is, that it offers a method for computing with words.

3. State of the Art

Some work has been done in the field of modeling and using physiological data in different applications and scenarios. Some of the most important related work and approaches will be described in this chapter to narrow down the contribution of this work. At first, the current research areas addressing this topic will be introduced. In a second step, different approaches for continuous modelling of physiological data to emotions or other states will be introduced and a conclusion will be drawn. In the following chapters, different approaches in the area of using mental load or affective state in applications or user interfaces as well as context-sensitive applications will be described. The chapter finishes with a conclusion on the current state-of-the-art of combining physiological signals with context information for applications.

3.1. Research Areas

Different research areas exist, that are based on the usage of physiological signals as input. The two most commonly used terms are ”Affective Computing” and ”Physiological Computing”. In the following a short definition of both areas and their challenges will be presented.

3.1.1. Physiological Computing

One of the current research areas dealing with physiological signals in applications is the area of Physiological Computing. Fairclough [Fai09] defined physiological computing as:

> ”Physiological Computing uses real-time psychophysiology to represent the internal state of the user (e.g. cognitions, motivation, emotion), which is used as the basis for real-time system adaptation.”

Physiological Computing focuses on psychological states based on physiology. The biocybernetic loop is the heart of a psychophysiological computing system [Fai09]. The loop includes the whole cycle of measuring and interpreting data up to the adaptation and the reaction of the user. Figure 3.1 shows the cycle of a biocybernetic loop based on [PBB95].

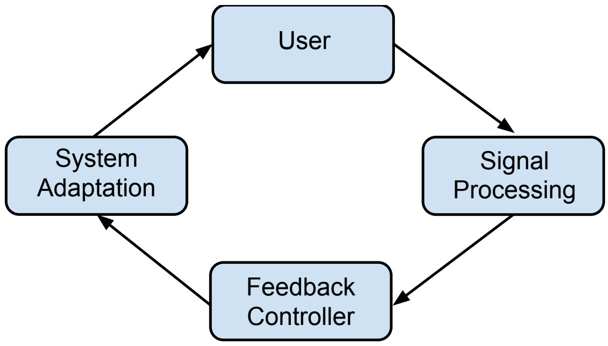


Figure 3.1.: biocybernetic feedback loop based on [Fai09]

The feedback loop is able to work in a positive or negative way for feedback control or switch between both. Negative feedback control aims on keeping the user in a certain state. The positive feedback control on the other hand aims for performance instability, for example to increase the challenge or engagement of a user. Depending on the application area, either positive or negative feedback control supports the purpose of the application. Games e.g. may benefit from positive feedback control. Keeping an user in a certain work-load zone on the other hand may benefit from negative feedback control.

3.1.2. Affective Computing

The area of affective computing was introduced by Rosalind Picard, who defined in 1995 in [Pic95] Affective Computing as follows:

> ”I call ”affective computing”, computing that relates to, arises from, or influences emotions.”

In contrast to Physiological Computing, affective computing focuses on the emotional or affective aspect of a user or the system, where physiological computing covers a broader range of different user states. Affective Computing has the aim to give computers and machines the possibility to handle affective state of an user, based on input from different sensors [TT05]. Besides covert physiological signals as used in physiological computing, affective computing also makes use of measures like facial expression and speech.

The current challenges in affective computing are similar to the challenges in physiological computing. In both areas, reliable models for user state have to be explored more in detail.

3.2. Continuous Modelling of User State

Different models and approaches exist, to interpret physiological signals. One of the challenges is to assess affective or emotional state continuously during runtime of applications. In chapter 2.9 different models were introduced, which are the base for the specific state-of-the-art models that are used in current research. These current efforts in research, covering different methods, will be presented in this chapter. Especially the type of affect or user state and its values, which are determined, is of interest for this work as well as the applied method itself and its classification rates.

3.2.1. Neural networks

Approaches based on neural networks are widely distributed. Nicolaou et al. [NGP11] developed an approach using Long Short Term Memory Neural Networks (LSTM). The methodology of the approach is shown in figure 3.2. Input signals are facial expression, shoulder gesture and audio cues.

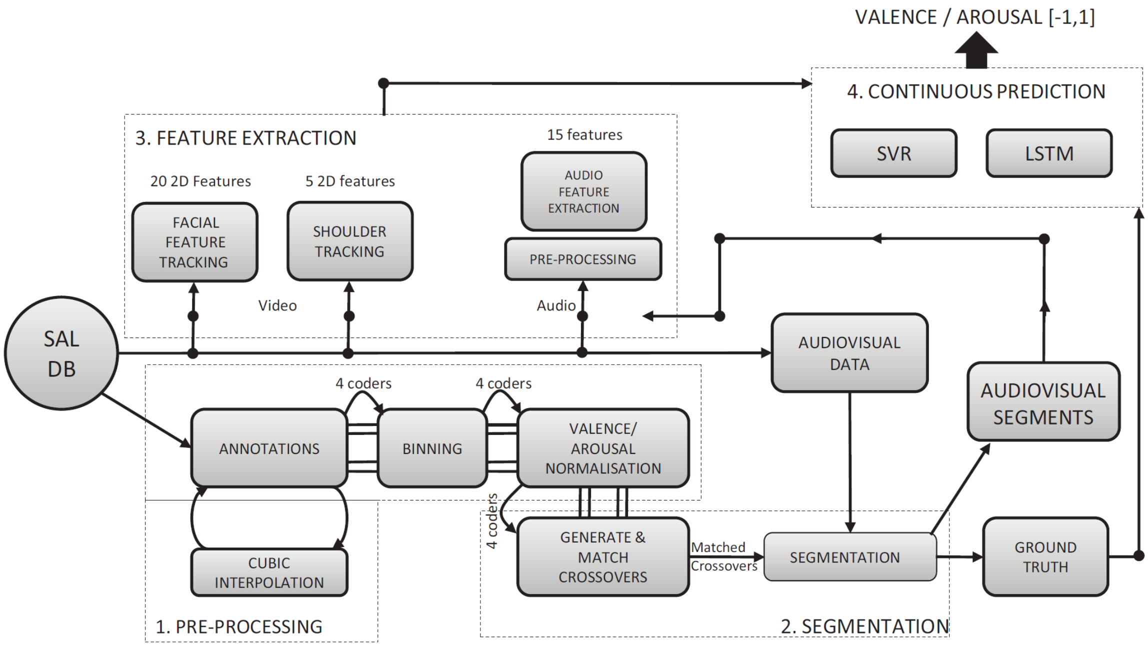


Figure 3.2.: Methodology and model of Nicolaou et al. [NGP11]

The approach is divided in four steps: Pre-processing, Segmentation, feature extraction and continuous prediction. The Sensitive Artificial Listener Database (SAL-DB) is the basis for pre-processing and segmentation, as well as feature extraction. The prediction part of the model is based on two different approaches: a variation of neural networks, the LSTM neural network and the Support-Vector Regression (SVR).

LSTMs have different advantages in comparison to recurrent neural networks (RNN) for the continuously modelling of affect and emotional states. LSTMs and RNN differ by the nodes, as the LSTM has connected memory blocks as nodes. These memory blocks contain a storage and multiplicative gates. The gates handle the state of the memory cell. This allows the LSTM to learn over a longer time than RNN.[NGP11]

Evaluated against a human coder, the prediction of valence and arousal values were similar or slightly better, than an average human-intercoder. In single cue-prediction, the system achieved a correct prediction of 84% for valence. [NGP11] LSTMs have been used mainly in models that were based on speech and facial expression recognition. Especially for speech recognition, the bidirectional version of LSTMs seems to be a good solution [GS05].

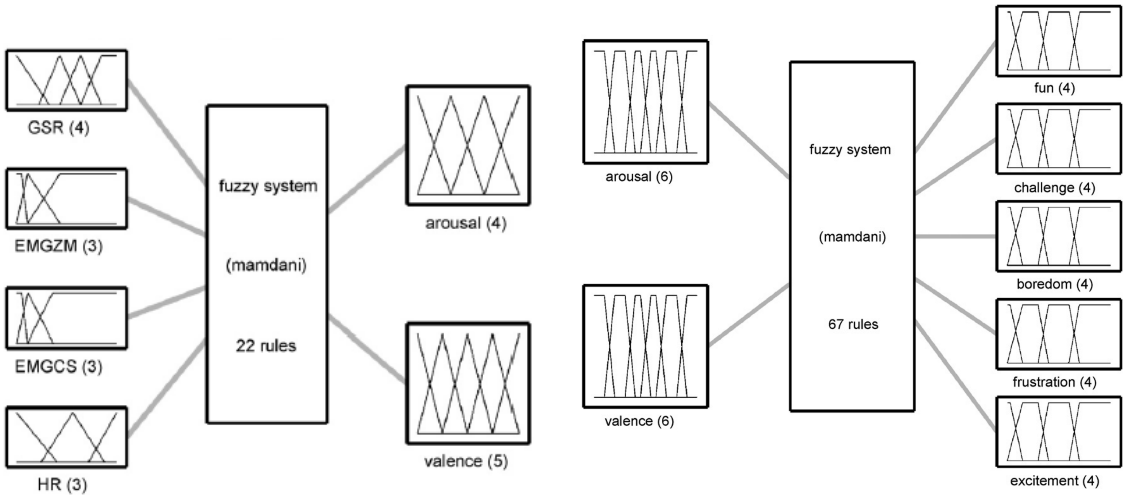
3.2.2. Fuzzy Logic

Several examples of current research exist, that use an fuzzy logic approach for interpretation of affective state. Mandryk et al. presented in [MA07] an approach to model physiological data continuously to emotional state. Based on the affect grid of Russell presented in chapter 2.7.2, the created model is specialized to computer games, modelling the states fun, boredom, challenge, excitement and frustration.

Mandryk et al. used Galvanic Skin Response (GSR), electrocardiogram (ECG) and electromyogram (EMG) to compute emotional states. ECG has been measured by three electrodes placed on the chest. For EMG measurement two electrodes have been placed in the face of the participants to measure smiling and frowning activity. Electrodermal activity (EDA) was measured by two electrodes placed at two fingers. These physiological signals get normalized in a first step. Based on these normalized values, values for valence and arousal are calculated (figure 3.3 (a)). The valence and arousal values get transformed to an emotion value.

To transform the physiological signals into valence and arousal values 22 rules based on fuzzy logic are applied to the normalized signals (shown in figure 3.3(a)). Arousal is based on EDA and heart rate (HR). On the other hand valence is generated by the two EMG values and heart rate. The 22 rules are based on the relationship between physiological signals and psychological interpretation and are presented in detail in [MA07]. Valence and arousal are expressed in six different levels: very low, low, mid low, mid high, high and very high.

In the second step (shown in figure 3.3(b)), valence and arousal are transformed into emotion values by 67 rules. The rules are based on the affect grid and can be found in detail in [MA07]. The affective grid (see 2.7.2) was modified, the nine point scale of the grid was reduced to a six point scale. The areas for the five different emotions in the affective grid, shown in figure 3.4 were defined based on the circumplex model of Russell. The x- and



(a) Transformation of physiological signals to valence-arousal space

(b) Transformation of arousal and valence to five affective states

Figure 3.3.: Model of the fuzzy-rule system by Mandryk et al. [MA07]

y-axis divide the grid in the vertical and horizontal direction. As in the affective grid, the x-axis ranges from unpleasent to pleasant feeling and the y-axis ranges from sleepiness to high arousal. The output of the emotion was divided into four levels: very low, low, medium and high (figure 3.4).

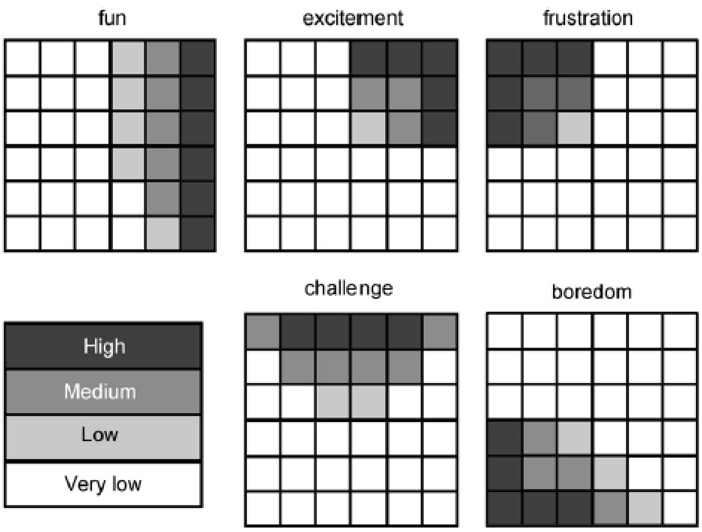


Figure 3.4.: The five affective output states in the affective grid [MA07]

The model was created with data sets of six participants. Six other participants were used to evaluate this model. Besides measurement of physiological data, participants had to fill out subjective ratings about their current emotional state. The results showed a correlation for fun and excitement. Challenge correlated with one half of the participants. The others commented, that they calm down during high degrees of challenge, which does not meet the model of challenge. Frustration and boredom did not correlate with the subjective ratings. Mandryk and Atkins state that scaling issues might be the problem in this case.

Another example for the usage of fuzzy logic is the work of Rani et al. [RSSA03]. The work of Rani et al. focuses on detecting anxiety instead of a broader range of affective states. Cardiovascular Activity, EDA, EMG and temperature were measured. From the 18 measured signals, only signals with a high correlation rate were used as input for the fuzzy system. The results of a first study were used to formulate the rules, which resulted in 3x n rules for n input channels.

In a first study of Rani et al. [RSSA03], six participants had to solve three different tasks with two different difficulty levels each to produce different levels of anxiety. Participants had to rate their subjective anxiety level after each task on several 9-item Likert-scales. The output of the anxiety classifier system was compared to the subjective ratings. The results showed a mean percentage error between 14 to 16 percent for different training data sets. In this study, the fuzzy logic approach was also compared to a decision tree approach. The mean percentage error for the decision tree ranged between 9 and 41 percent for the same data sets. The decision tree approach performed in best case better than the fuzzy logic approach, but the fuzzy approach turned out to be more stable across different situations.

3.2.3. Support Vector Machine

The support vector machine (SVM) is in comparison to the other used machine learning methods the newest method. Several current research uses support vector machines. Sun et al. [SKC+12] used a SVM for interpretation of stress based on ECG, EDA and an accelerometer. For evaluation, data set of 20 participants in six different situations were used. With a randomly chosen subset of the gathered data sets, the SVM was trained and afterwards evaluated with the remaining data sets. The classification rate was based on the training sets size. With only 3 out of 20 data sets, a classification rate of approximately 73 % was achieved. Using 18 out of 20 data sets, led to a classification rate of 81%.

Sun et al. [SKC+12] compared the SVM approach to Bayes networks and a decision tree. Bayes networks had the second best classification rate, ranging from 67% for 3 training sets to 78% for 18 training sets. Decision trees had the worst result in comparison to SVM and Bayes Networks, ranging from 64% to 78% classification rate.

Zhai et al. [ZB06] used a SVM to distinguish between stressed and relaxed states, based on EDA, ECG, skin temperature and pupil diameter. In total, 32 participants took part in the study. Participants had to do the Stroop Color Stress Test [Str35] in different configurations. Out of the 32 data sets, 12 were used for training and the remaining 20 for evaluation. Classification rate was on average 90.1%.

Liu et al. [LRS05] used SVM in a study to determine anxiety, engagement, boredom, frustration and anger in real-time. EDA, ECG and EMG were used as physiological input. The participants had to solve two different tasks: the game Pong and an anagram task. Both tasks were designed to cover a broad spectrum of the five selected affective states.

In total, 15 participants took part in the study. The SVM reached 85.8 % classification rate in average, ranging between 82.8 to 88.8% classification rate for single affective states. The SVM approach was compared to regression trees, Bayesian networks and k-nearest-neighbor approaches, with the same data sets. Judging by classification rate, SVM performed best, followed by regression trees, which had an average classification rate of 83.5%. K-nearest-neighbor and Bayesian networks had classification rates of 75.1 respectively 74%.

When comparing training and testing times, regression trees performed best. The SVM approach was two times slower than regression trees in training and 3 times slower in testing.

3.2.4. Other Approaches

Besides the mentioned approaches (neural networks, fuzzy logic and support vector machines), several other methods have been used in research. Some of them have been mentioned in the previous subchapters, as they have been compared to one of the other methods in studies. One other approach, which was one of the first ones with a high classification rate, will be described in the following.

In the work of Picard [PVH01] et al. eight different affective states, shown in figure 3.5, were defined, based on arousal and valence. To determine valence and arousal, EMG, EDA, blood volume pulse and respiration were measured.

Emotion	Imagery	Description	Arousal	Valence
(N)o emotion	blank paper, typewriter	boredom, vacancy	low	neut.
(A)nger	people who arouse rage	desire to fight	very high	very neg.
(H)ate	injustice, cruelty	passive anger	low	neg.
(G)rief	deformed child, loss of mother	loss, sadness	high	neg.
(P)latonic love	family, summer	happiness, peace	low	pos.
Romantic (L)ove	romantic encounters	excitement, lust	very high	pos.
(J)oy	The music "Ode to Joy"	uplifting happiness	med. high	pos.
(R)everence	church, prayer	calm, peace	very low	neut.

Figure 3.5.: Table of 8 different emotional states used in the work of Picard et al. [PVH01]

The model based on a hybrid sequential floating forward search with Fisher projection (SFFS-FP). It was compared with a solely sequential floating forward search (SFFS) and Fisher projection (FP) in the study. The results showed significantly better results for the SFFS-FP with a classification rate of 80.8 % for the 8 different affective states.

3.2.5. Conclusion

The different approaches presented in this subchapter have different advantages and disadvantages for usage in a big variety of applications and in mobile scenarios. None of the models reaches a 100% classification rate. Different aspects have to be considered when choosing a model for affect classification. One aspect is the number of input and output channels. A model, that works well for distinguishing between only two affective states based on one input signal, might not achieve the same high classification rates for multimodal input and a higher number of affective states output.

When favoring a model with a classification rate as high as possible, the support vector machine approach seems very promising. Different comparative studies showed, that the SVM approach had the best classification rates under different conditions. On the other hand, when aiming for a fast and efficient training and testing process, regression trees outperformed the SVM.

Besides SVM, fuzzy logic also achieved high values in classification rate in the presented work. In comparison to SVM, fuzzy logic has its advantages in the comprehensibility of the configuration. Rules can be defined with words, which are also understandable for persons without computer science background.

3.3. Integrating Workload Measurement in Applications

Various works did research about the usage of heart rate variability and other physiological signals in user interfaces, especially regarding reducing complexity or multi-tasking scenarios. Rowe et al. [RSI98] did a study comparing mental effort determined by Heart Rate Variability (HRV) and subjective measures. Chen and Vertegaal [CV04] implemented a user interface integrating mental workload for interruption management. Afergan et al. [APS$^+$14] based their work on a brain imaging technique. The examples will be further discussed in the following.

3.3.1. Difficulty Adaptation in Air Traffic

In the work of Rowe et al. [RSI98] participants of a study had to control an air traffic scenario in a game, preventing objects to collide. The parameter movement was divided into constrained and unconstrained movement, adapted in the aspects of speed, directions and moderation. Objects with constrained movement had predefined routes.

Thirteen persons participated in an early study. Participants were connected with three electrodes to a stationary ECG. Each participant had to play five different scenarios with different difficulty levels. For subjective rating of mental effort, NASA Task Load Index (NASA-TLX) was used. The results showed a decrease of HRV in the low frequency area which means an increase in mental effort, when task difficulty was increasing. This correlated to the subjective rating, which also showed an increase in mental effort.

When task difficulty reached a certain difficulty level, HRV increased, which would mean a lower mental effort. Rowe et al. [RSI98] state that evidence suggested that participants reached the point where they moved from a resource-limited condition to a data-limited condition. Before reaching this level, participants increased their level of mental effort to hold a certain level of performance. After reaching the point, where participants got more data, than they could handle and realize, that a certain performance level can not be hold due to limited resources. In this case, participants seemed to accept a lower performance and lowered their mental effort.

3.3.2. Dynamic Difficulty based on Brain Metrics of Workload

Similar to the study presented by [RSI98] the work of Afergan et al. [APS+14] used an interface for unmanned aerial vehicles (UAV) to test dynamic difficulty based on workload. Workload is measured with help of brain metrics based on functional near infrafred spectroscopy (FNIRS). FNIRS sensors require two probes placed on the forehead of a user, one on each side.

The interface used for the study is shown in figure 3.6. Participants of the study had to direct UAVs to targets (both shown in red). Several obstacles (shown in blue and yellow) had to be avoided on the track. When one of the UAVs arrived at one of the targets, new targets were shown.

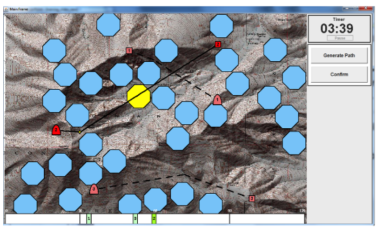


Figure 3.6.: Screenshot of the UAV test application [APS+14]

After a calibration task, to distinguish between high and low difficulty, participants had to do the UAV task in each condition for 10 minutes. The adaptive condition used an interface, which adapted count of UAVs on the map based on current workload. The non-adaptive condition on the other hand changed the count of UAVs randomly every 20 to 40 seconds.

In total, 12 participants took part in the study. Measurements like number of successes, failures and performance were used to compare both conditions. Results showed, that participants had nearly the same count of UAVs in both conditions (4.41 in the adaptive version to 4.69 in the non adaptive condition). There was no significant difference in number of completed tasks between conditions. The failure rate on the other hand is decreased significantly in the adaptive version in comparison to the non-adaptive version. The authors conclude, that participants are more attentive and engaged in the adaptive version, which leads to a lower failure rate.

Aferang et al. [APS+14] show in their work, that dynamic difficulty adaptation can lower the failure rate in an application. Studies and application were conceptualized for a stationary environment.

3.3.3. Physiologically Attentive User Interface for Interruption Management

Chen et al. [CV04] developed a mobile application, which uses heart rate variability and motor activity based on electroencephalography (EEG) to regulate notifications. The authors distinguished between four different states shown in figure 3.7. States were divided by high and low mental workload as well as high and low motor activity. When mental load and motor activity were high, it was assumed that user were e.g. writing, meeting or lecturing. If mental load was high, but motor activity low, possible activities of the user are driving, reading or thinking.

	Low Motor Activity (EEG)	**High Motor Activity (EEG)**
Low LF Power (ECG)	*User State 1* -Low mental activity -At rest *Candidate Activities* Pausing, Relaxation.	*User State 2* -Low mental activity -Sustained movement *Candidate Activities* Moving.
High LF Power (ECG)	*User State 3* -High mental load -At rest *Candidate Activities* Driving, Reading, Thinking	*User State 4* - High mental load - Sustained movement *Candidate Activities* Meeting, Lecturing, Writing

Figure 3.7.: Table showing the definition of the four different user states [CV04]

Based on these states phone call, messaging and email notifications were adapted as well as the messaging status. In state 1 phone call notifications are set to ring, messaging and email notification to vibrate and messaging status to available. State 2 differs only slightly from the adaptations of state 1. The profile is the same, except the messaging status, which is set to busy. In state 3 all notifications are set to vibrate and messaging status to available. In state 4, the state with high mental load and high motor activity, all notifications are set to silent mode and messaging status to busy.

A wearable system was used for measurement of physiological data, which allows a live transmission of the physiological data. For measurement of ECG three electrodes were placed at the chest. EEG was measured with one electrode placed on the head of the user.

In a first user study the correct notification state was identified in 83 percent of the cases in a group of 6 participants. This work presented by Chen and Vertegaal is one of the first systems, which was considering the usage of physiological data in mobile scenarios. Sensors were chosen to be wearable but needed wires to connect the electrodes to the measurement system. One of the electrodes was fixed at the head, which might lead to a low user acceptance in terms of usability.

3.4. Integrating Affective State in Games and Applications

One big research area on the usage of physiological data is the gaming area. Several models and games have been developed and researched, using physiological data as a direct or indirect input for gaming allowing to adapt or control a game. Several examples will be introduced in the following.

3.4.1. Integration of Physiological Signals in Gameplay

Drachen et al. [DNYP10] describe a study with three different first-person shooter games. Electrodermal Activity and heart rate were measured during gameplay and compared to subjective ratings of the gamer. The subjective rating was measured with game experiencequestionnaire (GEQ) by Ijsselsteijn et al. [IPDK08].

In total 16 participants took part in the study. The commercial games Prey, Doom 3 and Bioshock were tested. All of these are first person shooters with horror elements. Each participant had to play each game 20 minutes and was interrupted every 5 minutes to fill out the game experience questionnaire. Heart rate and EDA was measured with a stationary device. Physiological signals were normalized using the average baseline over all measurements.

The results showed correlations between the measured arousal level and the results from the subjective ratings. Heart rate correlated negatively with the GEQ dimensions competence, immersion, flow, challenge and positive affect. Heart rate correlated positively to tension and negative affect. [DNYP10] conclude that this is an indicator that a high heart rate is linked to frustration and tension, where on the other hand a low heart rate is linked to the flow state. EDA only correlated to negative affect and frustration.

Besides the work of Drachen et al., the research of Dekker and Champion [DC07] integrated physiological signals into the engine of Half-Life 2. For measurement, a Lightstone device from WildDivine was used, measuring ECG and EDA. Averages values of participants were gathered in previous measurements. The setup is shown in figure 3.8 (a).

(a) Game setup with physiological measurement device [DC07]

(b) Screenshot of the game [DC07]

Figure 3.8.: Biofeedback game of Dekker et al. [DC07]

The game was adapted in different ways. Movement speed and sound volume were adapted by heart rate and EDA. Other effects appeared if a certain threshold for one of the signals was reached. For example, the screen was shaking, when a certain high heart rate level was reached in comparison to the measured average values. Calming down lead to several advantages for the player, like semi transparent walls. Several visual effects were also adapted by heart rate, like the black and white version of the game (shown in figure 3.8 (b)), which appeared if the heart rate was below the average. If a player was too calm, game AI adapted difficulty, by spawning of additional enemies.

In a first evaluation, 14 subjects took part. In total, 8 of 14 participants noticed the enhanced version in comparison to the normal version and 9 out of 14 prefer the enhanced version. 10 of the 14 participants recognized the influence of their physiological signals on visualization and 12 recognized the influence on other events (e.g. sound adaptation).

3.4.2. Using Physiological Data as Direct and Indirect Input

Nacke et al. [NKLM11] focused in their work on the classification of using physiological signals as direct and indirect input in games. A 2D jump and run game was developed (shown in figure 3.9), which is used with different physiological controls for a study.

Figure 3.9.: Screenshot of the game used in the study of Nacke et al. [NKLM11]

For implementation of the game a C library was implemented, called SensorLib, which offers an interface for different physiological sensors. For their research described in [NKLM11] blood volume pulse (BVP), Galvanic Skin Response (GSR), ECG, EMG, respiratory (RESP) and temperature (TEMP) sensor were integrated. For passing the data to the game, the physiological signals were processed.

Five different game mechanics were implemented, which can be controlled by physiological signals. One game mechanic is the enemy target size. As bigger enemies are the easier to hit, size of the shadow around the enemy can be increased, which also increases the hit range. Besides target size, the flame length of the flamethrower can be controlled, speed and jump height as well as the weather condition and boss speed in the boss fight. Besides these game mechanics, one more game mechanic especially for gaze movement had been implemented, Medusa's gaze. This ability allows the gamer to activate the ability by a special item and freeze/slow down enemies by looking at them during the ability is active.

In the study, participants had to play three different versions of the game. Two versions with physiological control and one control version without physiological integration. Sensor mapping and threshold values for game mechanics were collected in another study previous to this study. Participants had to play each game condition at least 10 minutes or until completion of the level.

Mechanic	Cond. 1	Cond. 2
Target size	RESP	GSR
Speed/jump	EKG	EMG
Weather/boss	TEMP	EKG
Flamethrower	GSR	RESP
Avatar control	Gamepad	Gamepad
Medusa's Gaze	Gaze	Gaze

Figure 3.10.: Sensor mapping for the first study [NKLM11]

Figure 3.10 shows the mapping of physiological signals and game mechanics in the two game versions with physiological input. RESP, TEMP and EMG were used as direct control. GSR and ECG on the other hand were used as indirect game control. Gaze tracking was a special case, which was in both versions.

Results showed that participants 8 of 10 participants prefer direct control (EMG, RESP, TEMP) in comparison to indirect control (ECG, GSR). Participants criticized the slow reaction time of indirect control measurements as there was no direct feedback to an action. In comparison to the game version without physiological signals, participants reported a greater immersion in the game versions with physiological control. The authors recommend to use indirect physiological input as dramatic device in games to alter the game world, when used for direct control.

3.4.3. Continuous Evaluation of Emotional Experience in Games

Mandryk et al. [MAI06] conducted a study, based on the model described in chapter 3.2.2. The goal of this work was to develop a methodology to measure playability and user experience in an objective way. The game NHL 2003 was used on a PlayStation 2 in the experiment. Physiological data was collected by a ProComp Infiniti system, which measured GSR, ECG, EMG for smiling and frowning. The game was played in three different situations: against a computer, against a stranger and against a friend. Before each session, players had to rest for five minutes.

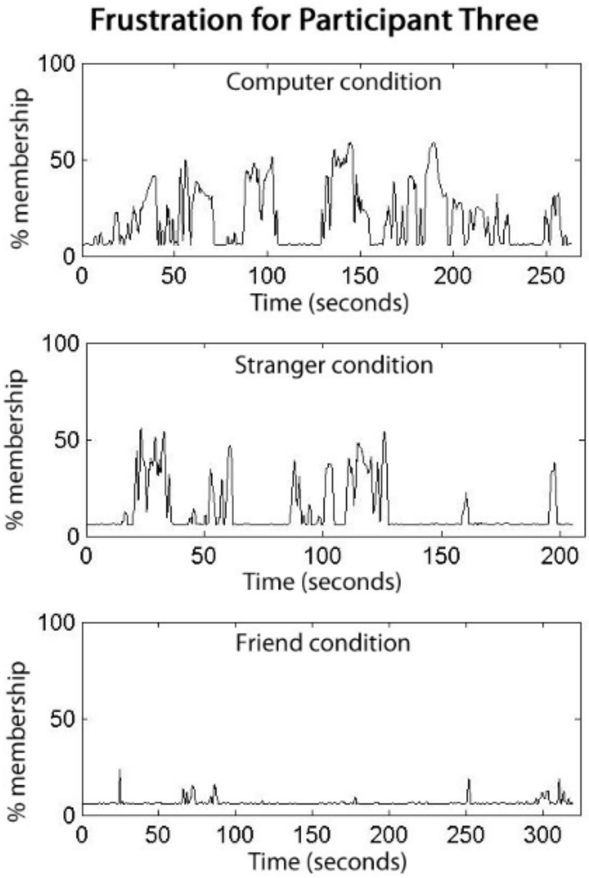


Figure 3.11.: Frustration of a participant under different situations [MAI06]

The physiological data sets were analyzed with the fuzzy logic approach of Mandryk. The results showed that participants enjoyed playing against a friend more, than playing against a stranger or especially a computer. Figure 3.11 shows the frustration of a participant under different conditions. The graphs show a significantly lower frustration during a friend condition in contrast to the computer or stranger condition.

Other work in the area of game evaluation using physiological signals has been done by Nacke et al. [NGL10] [NL10]. In comparison to Mandryk et al. [MAI06], Nacke et al. used the physiological signals without a direct mapping to affective states.

3.4.4. Influencing the Affective Experience - the Emotion Engine Framework

The work of Nogueira et al. [NRON13] addresses the aspect of immersion of gamer. Nogueira et al. developed a framework which learns based on the emotional state during different situations to allow modeling the affective gaming situation of a player.

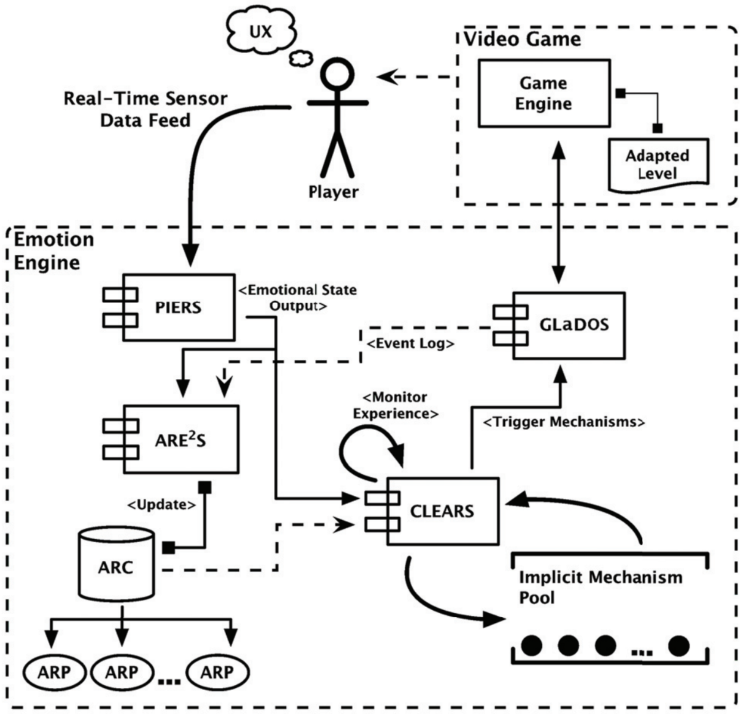


Figure 3.12.: Architecture of the Emotion Engine [NRON13]

Nogueira et al. [NRON13] propose the Emotion Engine biofeedback loop system. The framework consists of different separate components. One component, called PIERS interprets the physiological input in an emotional state. The following two components, ARES and CLEARS create an affective response profile and regulate based on the preferences of the user. The GLaDOS component acquires the occurring events in the game and passes them to the ARES component to adapt the affective response profile. CLEARS selects a game event or parameter, which is passed via GLaDOS to the game engine, when the emotional state of the user differs from the desired state.

As physiological measures heart rate, EDA and EMG are used and transformed into arousal and valence. The survival horror game Vanish was used to demonstrate a first integration of the framework in a game.

3.5. Conclusion on Integration of Physiological Signals in Applications

In chapter 3.3 several examples using mental workload and cognitive aspects, for adapting interfaces or applications, were presented. The shown examples all realized real time adaptation, but lack in addressing mobile situations. In many cases, sensors were used which might lead to impairment of the user during usage in mobile scenarios due to wires. Also influencing factors in mobile scenarios are not taken into account.

Besides cognitive aspects, different examples of current research focusing on usage of affective state in one or another way, were presented. A broad variety of similar examples exist, therefore this analysis outlined different aspects of integration of physiological data by a few examples. The presented examples showed, that user state, may it be affective or cognitive can be determined with relatively high classification rates. In most cases, the measured signals were analyzed after the study, instead of adaptation in real time. The work of Nogueira et al. [NRON13] presented an approach of an framework for real-time adaptation in games.

Most of the examples, except the work of Chen and Vertegaal [CV04], were developed with stationary measurement devices. The work of Chen uses mobile sensors, but does not address the mobile aspect and its challenges.

The introduced example of Mandryk et al. [MA07] in chapter 3.4.3 presents an approach of using affective state interpretation for evaluation in game research. The goal of this work was to evaluate the game but the approaches presented have the potential in use of real-time adaptation in applications.

In difference to the other research presented in this chapter, the aim of the work of Nacke et al., was not to classify emotional or affective states. But the work of Nacke proved an important point, that indirect used physiological signals may not get consciously perceived by the user.

The different research done shows promising models and concepts for interpreting cognitive or affective state of a user. But up to today, the analysis shows, that several aspects like mobility or real-time adaptation have only been addressed at the surface.

3.6. Context-Sensitive Interfaces

Context sensitive interfaces are able to react to the environment or a situation. Many examples for context sensitive interfaces exist as well in desktop as in mobile devices. Modern mobile devices offer a great variety of different sensors, an example is shown in figure 3.13. In the area of mobile context-sensitive interfaces, personal assistants reached a wide distribution amongst mobile devices within recent years.

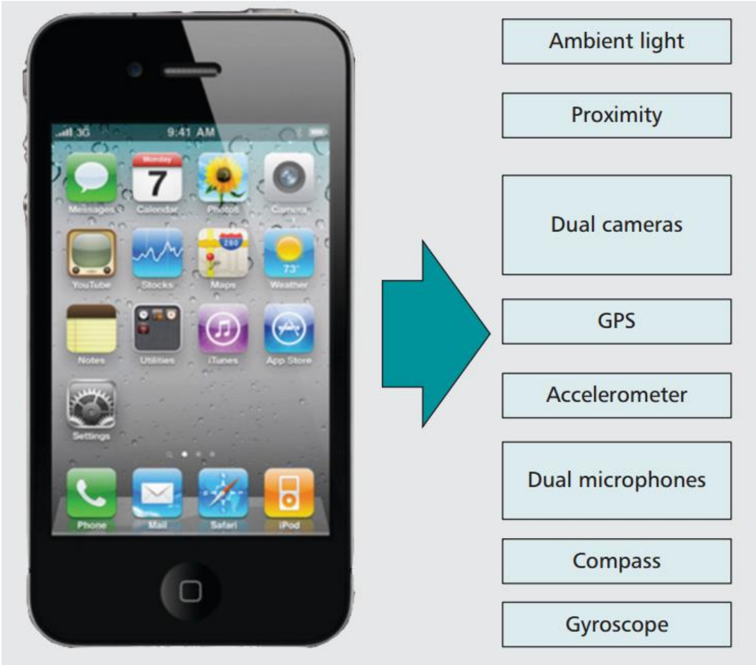


Figure 3.13.: Integrated sensors in the iPhone 4 [LML+10]

Mobile personal assistants manage information flow and offer several functionalities in one place for the user of a mobile device based on the current situation. Several examples exist, like e.g. Google Now, Microsoft Cortana and Apples Siri. All three mentioned examples do not use physiological signals but a broad variety of different kinds of context information as an input to arrange the overview of information [Inc15a] [Cor15b] [Inc15b].

3.6.1. Apple Siri

Siri is developed by Apple and was first released in September 2011 with iOS5. Currently Siri only runs on iOS devices. Siri is mainly a speech recognition software for assistance. Like Google Now and Cortana, Siri also offers a natural language interface. It uses information about past searches to predict and individualize future search results.

Functionalities that can be controlled via speech are for example setting alarms (figure 3.14(a)), writing messages (3.14(b)), showing information of contacts (figure 3.14(c)) or making calls. Speech recognition can also be used to write text in other applications of the device like e.g. applications for social networks.

(a) Setting of alarms

(b) Dictation of emails via voice

(c) Presentation of contact information

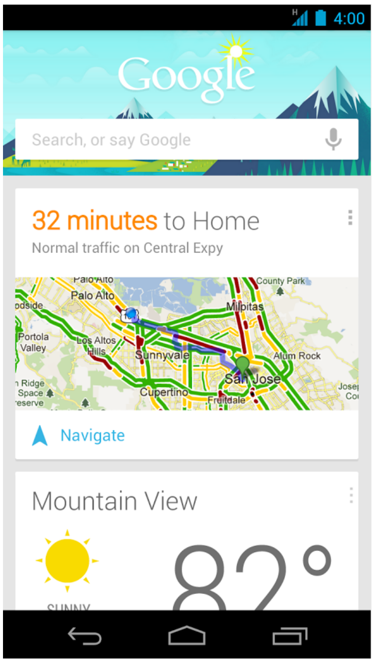
Figure 3.14.: Screenshots of Apple Siri [Inc15a]

Siri is able to react to requests by asking for details to refine and optimize results. In comparison to Google Now and Cortana, Siri uses less context information and offers less functionality.

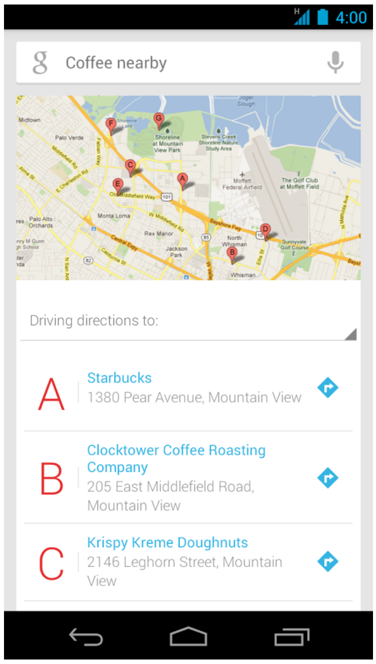
3.6.2. Google Now

Google Now is developed by Google and was released in 2012 with Android version 4.1. Currently Google Now is available for Android OS and Apple's iOS. It is mainly based on the Google search engine combined with several context information gathered from other applications and sensors of the device.

When opening Google Now, a screen with an overview of information that might currently be important to the user is shown. For example current weather forecast, traffic to work or nearby interesting locations are shown (for example as shown in figure 3.15 (a) and (b)). Google Now gathers information for example from the email inbox and the calendar. If an email contains information about a flight or a flight ticket, it will be shown at the day of the flight together with information about punctuality of the flight and when to leave the house to arrive on time at the airport. Regularly travelled routes are also shown, e.g. traffic for the way to the office in the morning and back home in the evening.



(a) Google Now screenshot with traffic and weather forecast



(b) Google now screenshots with results of coffee stores nearby the current location

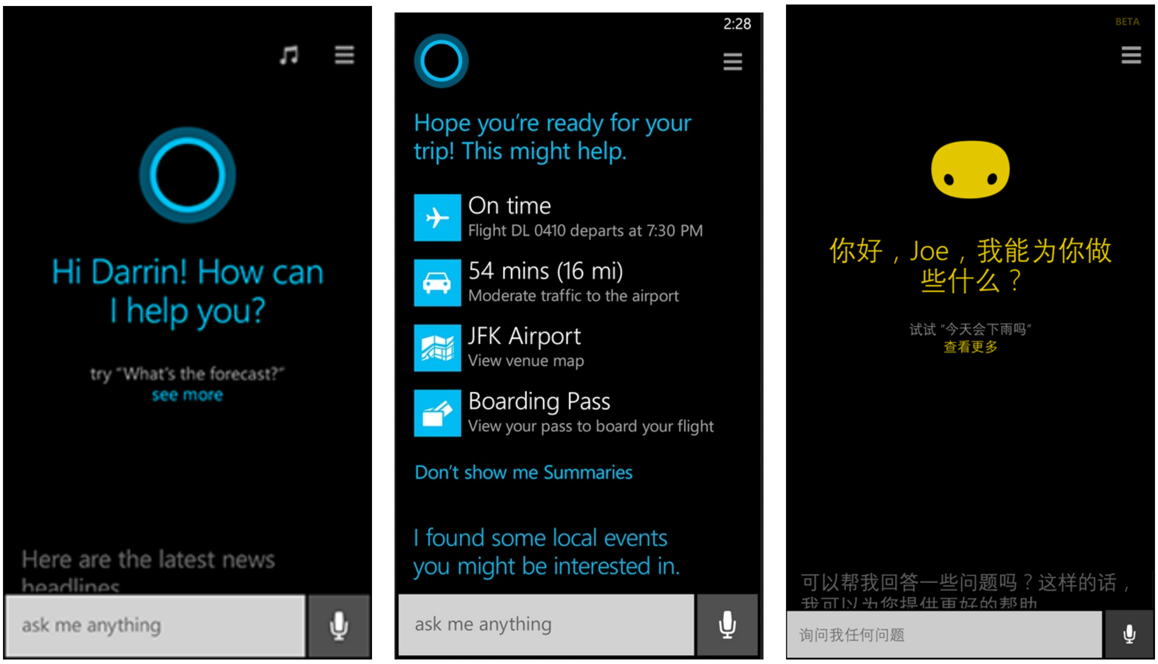
Figure 3.15.: Screenshots of Google Now [Inc15b]

As shown in figures 3.15, at the top of the screen, the Google search bar is shown, which can be used by typing an a item to search for or by saying the keyword ”OK Google” followed by the item of search. Besides the typical search function, Google Now supports different voice commands. For example an alarm can be created by the words ”OK Google set an alarm for .. ”. Besides alarm settings, other commands are available for e.g. information search, starting applications, navigating to an address, sending emails or searching for nearby stores. Answers to requests are given by showing of notifications or search results. When asked a question for search on the internet via voice recognition, Google Now reads the top answer depending on the results.

3.6.3. Microsoft Cortana

Cortana was developed by Microsoft and first released in 2014 with version 8.1 of Windows Phone. With the introduction of Windows 10, Cortana also became available for desktop systems, followed by Android and iOS versions in later 2015. The software is currently available in 9 different languages and 15 regions.

Cortana offers services similar to Google Now voice recognition for different activities, like setting an alarm or launching an application. The screenshots shown in figure 3.16 (a) and (b) show the start screen of Cortana and an overview for information around a flight. Features like traffic of frequently travelled routes and an overview of news tailored to the own interests are also offered. Other features are for example weather forecast, overview of appointments, sending messages and making calls. It is also built to adapt to regional differences, for example the voice feedback is talking in region specific idioms and the chinese version, known as Xiao Na, features a different set of graphics (c).



(a) Screenshot of the main view of Cortana offering voice recognition

(b) Information overview of a flight

(c) Xiao Na, the Chinese localisation of Cortana

Figure 3.16.: Screenshots of Microsoft Cortana [Cor15b]

A search bar is placed at the bottom and has similar functionality as the Google search bar in Google Now. Search queries can be done by typing in a term or asking Cortana via voice recognition. The answer is delivered by showing the search results or giving the answer to a question, if possible.

3.7. Conclusion of Context Information in Mobile Applications

Several state of the art examples of current research in the area of context sensitive user interfaces and applications have been introduced. The currently popular assistants allow adaptation to several aspects of context, offering information tailored to the situation, e.g. offering navigation or information about nearby places.

Up to date, integrated sensors in smartphones allow to conclude on many different context information of the environment of a user, like e.g. location. Only little can be said about the mental or psychological situation of the users themselves by only using context information that does not include physiological signals.

3.8. Combining Physiological Signals and Context Information

Only few examples exist, supporting or enriching the interpretation of user state based on physiological signals by context information. In the work of Sun et al. [SKC+12], context information is used, to control influencing factors on different physiological signals, to improve interpretation. Other research did take part outside the lab, but did not use context information for interpretation, like the work of Healey and Picard [HP05], who measured stress based on HRV during driving a car.

3.8.1. Activity Awareness

Sun et al. [SKC+12] addressed in their work the problem, that different physiological signals may react diverse when using outside of controlled laboratory conditions. In their work, ECG and EDA were measured. Especially heart rate is known for having different values in different conditions like sitting, standing or walking. To address the problem, Sun et al. used an accelerometer to detect movement.

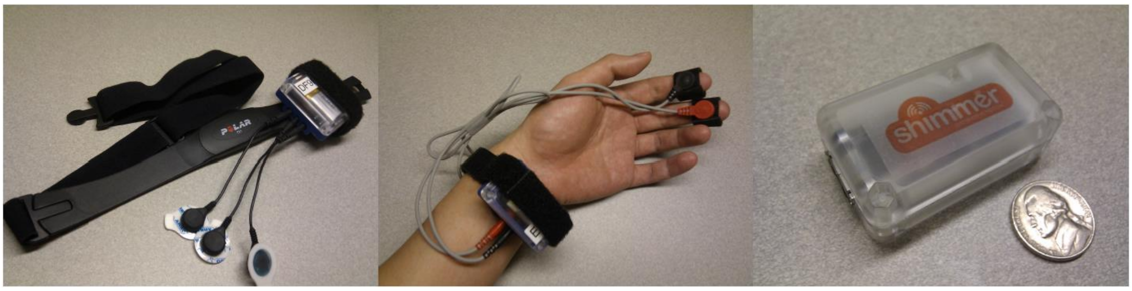


Figure 3.17.: Sensors for measurement of ECG, EDA and Activity used in the study of Sun et al. [SKC+12]

Figure 3.17 shows the sensors, which were used in the first study of Sun et al. Based on the position of the sensor, three different states of activity can be identified: sitting, standing and walking. Figure 3.18 shows the result of the first study.

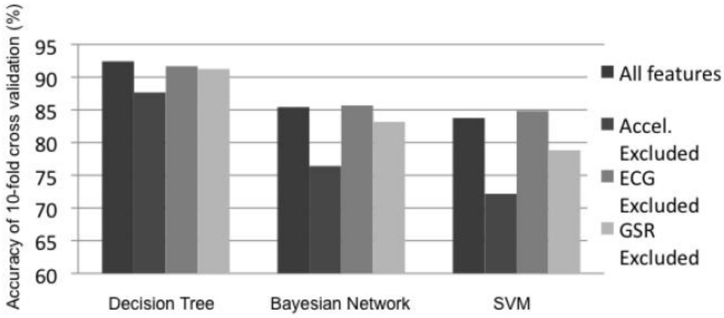


Figure 3.18.: Results of the study of Sun et al. [SKC+12] for the single subjects

The results showed, that accelerometer helped interpreting the mental workload of a person under different activities. The classification rate significantly decreased for the different classification methods, when accelerometer was not used.

3.9. Conclusion

Situations of smartphone usage can partially be identified with the help of context information. In chapter 3.6, several context-sensitive interfaces were introduced, that conclude on current situation the user is in, offering help and information. However, these interfaces do not give an insight into the state of the users themselves, like affective feelings or mental workload.

Different examples have been presented in chapter 3.3 and 3.4, using physiological signals to give an insight into the user. Physiological signals have been used to determine affective aspects, emotions and cognitive aspects like mental workload. The used models proved to deliver reliable results in the presented studies. On the other hand, the examples were designed for stationary settings. As many physiological signals are influenced by movement or other aspects, these examples may fail in mobile scenarios as the interpretation of physiological signals is influenced.

When combining the interpretation of physiological signals and context information for user state interpretation, the introduced examples of current research show, that there is still work needed. Many examples did not use context information in any way in their applications and examples to control influence of physiological signals by aspects like movement. Most evaluations presented in this chapter were only conducted in a laboratory with a controlled environment.

Besides the example of Sun et al. [SKC+12], which uses an accelerometer, plenty other context information exists, measured by sensors integrated in mobile phones. Examples for context-sensitive interfaces were presented in chapter 3.6, which show a broad band-width of usage scenarios for context information.

Already known models, presented in chapter 3.2, have to be examined for their suitability and usage in mobile scenarios, as well as their extension by context information. Further context information besides movement has to be examined for usefulness in supporting user state interpretation. The lack of context information in the user state classification and its suitability for mobile scenarios is part of this thesis, and will be presented in the next chapter.

4. Theses

This theses is divided into three theses to examine different aspects of physiological data and context information as an input for mobile applications. The integration of physiological signals allows to analyze the state of a user, which can be used to adapt an application to the users' needs. As described in the State of the art section, many integrations of physiological data into stationary applications have been done, but a lack of models for usage scenarios beyond stationary settings exist. One big challenge is the conceptualization and implementation of a model for mobile scenarios. This requires to examine if and how existing models for mainly stationary usage can be used and integrated in mobile scenarios.

In mobile scenarios many disturbing factors appear, which can at least be better controlled or do not occur in stationary scenarios. For example movement can influence physiological signals like heart rate. Furthermore, mobile applications have several limiting factors. They are currently limited e.g. in screen size and battery. Measures have to be taken, to guarantee the robustness and reliability of the model and interpretation of user state in different scenarios.

As a result of the requirements of the mobile scenario, a better knowledge of the current environment and situation of the user is needed. This requires to gather context information to integrate into the model for a better interpretation of user state.

In the following the theses of this work, investigating the mentioned problems and questions, will be introduced and in detail described in the following subchapters. Approaches for the solutions will be introduced within the following chapters.

4.1. Mobile Scenarios

Before defining the theses in detail, this section will give a brief overview of how a mobile scenario is defined within this work and how they differ from laboratory settings. Statistics showed, that mobile devices are used in many different situations. Situations of usage can differ by locations, movement or distraction of the environment. When a person is not at home or office, applications have to deal with typical limitations like battery lifetime and network coverage. Movement on the other hand can split the attention of an user, as the environment needs a certain amount of attention to prevent accidents. Many of these defined situations can be determined by context information.

4.2. Theses

Thesis I:

The combination of physiological signals and context information improves the interpretation of user state.

Mobile devices offer a broad range of sensors, which collect additional information about the context. We assume that context information helps to improve the interpretation of physiological signals and the model for interpretation itself by controlling influencing factors on physiological signals and providing additional information about the environment. This raises the following question:

- Can context information provided by application and mobile phone improve the interpretation quality of user state?

Chapter 4.3 gives a short overview how this thesis will be investigated in this work.

Thesis II:

A general model for the combination of physiological signals and context information as an input for user state interpretation in mobile applications can be defined.

The current state of the art in this area was described in the previous chapter. Only [SKC+12] investigated the usage of physiological data in mobile scenarios briefly. We assume that current models can be extended and modified to suit the needs of mobile scenarios. This raises the following questions:

- Can a model be defined that supports different types of applications, e.g. learning and entertainment?
- Can a model be defined, that supports the usage in mobile scenarios, e.g. usage of an application during travel without loss of quality in interpretation and without impairing the user?

A short overview, how the thesis and these two questions will be investigated is given in chapter 4.4.

Thesis III:

The model developed in thesis II is robust enough to handle the loss of input channels.

In mobile scenarios, there is always the possibility that one of the input channels gets lost because of e.g. battery life time or connection problems. We assume that the model is robust and reliable enough to compensate the temporary or permanent loss of one input channel and deliver reliable results. This thesis raises the following question:

- Can loss of a channel, e.g. by empty battery, be compensated without a big drop in interpretation quality?

Handling of channel loss and reliability in this work, will shortly be introduced in chapter 4.5.

4.3. Combination of Physiological Signals and Context Information

A difference between stationary and mobile scenarios is the broad range of possible environments and situations, which can occur and influence the behavior of an user. Modern mobile devices offer sensors like step-sensor, Global Positioning System (GPS) and many more. These sensor offer valuable information of the context the user is interacting in.

An analysis of available sensor data will be done and available information categorized. The data gets preprocessed and is then transmitted to the different input channels to correct external influencing factors on physiological signals. Furthermore, context information itself is used to improve the interpretation of user state, because it can be used as an additional information itself instead of a solely usage to control influencing aspects on physiological signals.

Thesis I aims to integrate context information into the model defined in thesis II, allowing a more reliable interpretation of physiological signals.

4.4. Model for Mobile Applications

Existing approaches for the integration of physiological data into applications need to be examined on their suitability for mobile scenarios. For mobile scenarios, processing of data needs to be fast and continuously, as the situation might change within seconds.

In this thesis existing approaches for user state interpretation are analyzed, used and extended for a broad range of mobile applications. The characteristics of mobile scenarios and its influence on usage of physiological data are investigated. As an input, physiological signals need to be chosen based on the restrictions given by the mobile scenario. Sensors for measurement need to be as small and unobtrusive as possible. The signals need to be continuously processed to achieve a high reliability in the interpretation of user state.

The output of the model is the current state of the user, like e.g. affective or cognitive state, which allows mobile applications to adapt the interface, content or other components.

With thesis II, a model for integration of physiological data into mobile applications will be validated. This allows the adaptation of user interfaces or content in applications based on the current state of the user to achieve the goal of a higher user experience in comparison to applications without support of physiological data.

4.5. Reliability of Input Channel Handling

The mobile scenario requires a certain robustness of the model, to handle the loss or corruptness of input signals. The model needs to adjust if one of the physiological signals or parts of the context information is missing. With the validation of thesis III, the proposed model is adapted to handle the loss of input channels with an minimal loss in accuracy of user state interpretation.

To assure a robust model, the question, how big the accuracy of the model is, when one measurement or input channel falls away, has to is investigated to give an appropriate estimation for different scenarios. This information itself can be used as feedback within the model. The model needs to be tested with several configurations for different applications to give an estimation.

4.6. Conclusion

As introduced in thesis II, a model for integration of physiological signals into mobile applications is needed. Different physiological signals will be examined about their suitability for mobile usage. Existing models will be analyzed and adapted to fit the usage scenarios. As a result the model will deliver an estimation of current user state, which is defined by different categories. These values can be used as an input in mobile applications of different areas for adaptation.

The combination of physiological signals and context information gathered by integrated sensors of mobile devices leads to more accurate and reliable results. User status is refined and influences on the physiological signals controlled. Furthermore, the combination of both input channels leads to a higher robustness and reliability of the model.

5. Concept

After introducing the state of the art and theses of this work, this chapter will describe the overall concept and architecture of the developed model and will give an overview over different parts of the concept. In a first step the requirements for the use of physiological data in mobile scenarios are described. In the following, sensors and physiological signals are analyzed and examined regarding their suitability for the scenarios. Chapter 5.3 presents an analysis of different context information and the usage of this in the model.

The output of the model is defined, as well as a method is chosen for interpretation of the signals. Mobile applications that might use this model are addressed. Finally, the chapter closes with an overview of the biocybernetic loop and a presentation of the abstract model.

5.1. Requirements in Mobile Scenarios

Mobile scenarios differ in many aspects from solely desktop scenarios. They are characterized by a broad range of different situations that can occur during the usage of mobile applications. Users are not bound to a location, often multitasking is done during usage of applications (e.g. walking to the bus, driving etc.). The resources of a user are limited.[TOTK04] [OTRK05]

The state of a user can be influenced by different surrounding factors, like e.g. noise or traffic. Furthermore interaction with application can be interrupted by external factors, like interaction with other people. Therefore context has to be considered for a correct interpretation of user state.

Besides these aspects, the hardware on which the applications are running differs from desktop versions. Mobile devices offer only limited resources regarding screen space, battery lifetime etc. Devices are very heterogenic and offer a great variety of hardware and operating systems (e.g. iOS, Android, Blackberry).

User often have only sparse time to use an application when travelling or being at a public place. Therefore, the solution for the theses defined in chapter 4, needs to fulfill several criteria. In the following subsections, different aspects of the concept and their requirements will be introduced.

5.1.1. Measurement of Physiological Data

Measurement of physiological signals requires the user to wear different sensors. It might be useful, to have sensors that are comfortable to wear and allow free movement of the user in different situations that can occur in daily routine. Therefore transmission between sensors and mobile device needs to be wirelessly and have to allow continuously reliable measurement. Users should not feel impaired in any way or be constantly reminded of wearing sensors. One other important aspect is, that the measurement equipment should not look curious to not make the user feel intimidated.

Physiological signals have to be chosen based on these requirements. An overview of signals used in related works and which physiological signals will be used in this work and which sensors have been chosen will be explained section 5.2.

5.1.2. Processing of Physiological Data

As the signals and values of physiological data vary from person to person in most cases, baselines have to be measured and the processing has to be adapted to the user. The more data is measured, the better the accuracy of interpretation will be. Therefore it is preferred to measure the signals over a long time frame in the background instead of only for the timespan of application usage. Data needs to be processed effectively in the background to not use more resources of the already limited resources of a smartphone as needed.

This requires a data structure for collecting physiological data of a person, which is frequently reset when a certain time frame expires or sensors are not worn any longer. Reconfiguration of baseline and physiological sensors have to be considered.

5.1.3. Models for Classification

Several different models, which can be used to classify user state, have been introduced in the background and state of the art chapter. Most of them did not address the mobile aspect explicitly. To fit to mobile scenarios a model is needed, that only needs a short time for learning about the physiology of a user. Rather the model should learn in the background and only need a short period of time for baseline measurement and configuration.

To improve robustness, the model should support a fast reconfiguration phase as it is possible, that sensors are taken off by the user and not worn for some time. Depending on the measurement some sensors require to remeasure the baseline, as the values might vary in comparison to the earlier measurement.

Besides the integration of physiological signals, the model is also required to have a high flexibility for the integration of context information. Depending on the mobile device of an user not all sensors for context information might be available. The model needs to be robust and reliable enough for such situations.

5.1.4. Output

Applications on mobile devices have a great variety, ranging from applications like games to more performance oriented office applications. The state of the art analyses showed, that it is possible to determine different emotional, affective or cognitive aspects of an user from physiological signals. As the presented examples in the state-of-the art section in chapter 3 show, games and applications, which have the goal to entertain an user, may benefit more from the affective state instead of cognitive aspects. On the other hand applications, aiming for a certain performance level of the user and the goal to keep an user within this level, may benefit from information about cognitive aspects. Examples for such applications are learning and office applications or applications that need a high amount of mental resources.

Based on that analysis the concept will cover affective and cognitive aspects to allow support for a broad variety of applications.

5.2. Physiological Signals

For further thoughts about the concept and the model it is important to decide, which physiological signals are measurable in mobile scenarios. A broad range of signals is used in different applications, but not all of these are applicable for mobile scenarios. Requirements have been described in section 5.1.1. In this section, different physiological signals will be reflected briefly regarding these requirements. A decision will be made about the signals as well as sensors will be presented, which will be used.

5.2.1. Overview of Physiological Signals

Not all physiological signals can be measured without impairment of the user. For the use in mobile settings, it is important to use unobtrusive sensors that allow wireless transmission of data as described in the requirements section 5.1.1. The sensors need to be as small as possible to achieve a high user acceptance.

For that reason, the concept is based on electrodermal activity and heart rate. Both signals can be measured with relatively small sensors. As described in 2.5.1, electrodermal activity can be used to determine the arousal level of an user. Heart rate can be used to determine the valence in some cases or to run a spectral analysis for determination of mental effort (see chapter 2.4).

Other physiological measures like electroencephalography (EEG) or electromyogram (EMG) require at the moment relatively big sensors or sensors in a prominent place. For example EMG electrodes are commonly placed in the face. EMG is used in many research projects and applications, described in the State of the Art section in chapter 3. But most of these projects had not to deal with mobile scenarios and its requirements.

Due to these constraints, a heart rate monitor that allows Bluetooth transmission as used in sports is used for heart rate measurement (e.g. Polar or Zephyr HxM). For measurement of electrodermal activity, the Q-Sensor from Affectiva will be used as input channel in the model. In early studies, wired stationary devices were used for first evaluation of applications. In later stages, Q-Sensor, Zephyr HxM and Polar H6 were integrated.

Two heart rate sensors were integrated because of the different options and additional sensors they offer. Polar H6 offers a higher battery lifetime but otherwise Zephyr HxM offers more additional information measured by the heart rate monitor. Depending on the application scenario, one of the sensors might be more suitable than the other. Q-Sensor was chosen because it measures wirelessly and reliable Electrodermal activity (EDA).

5.2.2. Affectiva Q-Sensor 2.0

The Q-Sensor was originally developed at the Massachusetts Institute of Technology by Poh et al. [PSP10] and was distributed as Q-Sensor 2.0 curve by Affectiva. Besides skin conductance it measures temperature and movement. It allows wireless transmission of the measured data via an integrated Bluetooth module.

Figure 5.1.: Q-Sensor 2.0 from Affectiva [Aff11]

As shown in figure 5.1 the sensor is worn as a wristband. Two Ag/AgCl electrodes are placed at the inner side of the arm. As described in the background area, EDA is normally measured at the palm. A study described in [PSP10] shows that EDA can also be measured with the Q-Sensor at the distal forearm. Furthermore the participants of this study did not feel discomfort in a long time study. In this long time study of Poh et al. [PSP10] the Q-Sensor was also tested against devices, which are approved by the Food and Drug Administration (FDA) in the United States. Results showed, that the Q-Sensor offers very high accuracy comparable to FDA approved devices.

With help of the integrated sensor for movement of the arm and body temperature, influencing factors can be controlled in the interpretation of EDA values. Q-Sensor has been used in several other research projects, e.g. in the evaluation of child-robot interaction [LHMP13].

5.2.3. Zephyr HxM

Zephyr HxM (shown in figure 5.2) is a heart rate monitor consisting of a strap and a transmitter. Transmission is done wirelessly via Bluetooth. Out of the cardiovascular measures, Zephyr HxM measures heart rate and RR-intervals. Furthermore calories, steps, speed and distance can be measured with integrated sensors. Integrated algorithms address the problem of noise and movement artifacts [Tec10].

Zephyr HxM has been used in different research projects like in a phone-based health assistant [SBVL11], the involvement of audience at public events [PTS+10] and different research projects at the National Aeronautics and Space Administration (NASA) [RC12].

Figure 5.2.: Zephyr HxM BT [Tec10]

The data is transmitted wirelessly via Bluetooth 2.0. A version supporting the Bluetooth Smart standard was not available at the time of this work. Because of the Bluetooth 2.0 transmission, the battery life time is only about 26 hours. In comparison to Polar H6, battery can be recharged with an included charging station. An open software development kit is offered by Zephyr Technologies and the heart rate monitor is supported by different publicly available applications.

5.2.4. Polar H6 Heart Rate Monitor

The Polar H6 (shown in figure 5.3) heart rate monitor consists of an adjustable strap and a transmitter. The electrodes are on the inside of the strap. Strap and transmitter are both water resistant and have a battery lifetime of around 300 hours. The Polar H6 measures RR-intervals, heart rate (HR) and several other additional information.

[NDJ+09] proved the validity and reliability of former versions of the H6, the Polar

RS810 in comparison to electrocardiogram (ECG) devices. The results showed, that short term Heart Rate Variability (HRV) measurement is as reliable and valid as stationary devices. Several commercial applications exist, that support the Polar H6.

Figure 5.3.: Polar H6 [Ele13]

The H6 supports the Bluetooth smart standard, which allows the wireless transmission with low energy consumption leading to longer battery lifetimes. The transmission range is approximately 10 meters, which should be sufficient for most scenarios.

5.3. Context Information

Modern smartphones offer a variety of sensors, which can be used to collect context information as presented in the state of the art section 3.6. Based on a gyroscope or magnetometer sensor, activity of an user can be determined. For example, if an user is moving, how fast and in which direction. Integrated Global Positioning System (GPS) sensors offer location information.

These two measures are especially interesting for the model. A further distinction of how context information is categorized and used in the model is described in detail in the following subchapter. Some of context information that can be acquired is shortly introduced in the following, followed by a conclusion on the choice of integrated context information.

5.3.1. Categorization of Context Information

In a first step, context information is categorized depending on the usage within the model. Several categorization concepts for context information exist. For this work, context information has been divided into two main categories, the two main purposes of context information in the model: correction and improvement. Both categories are presented in figure 5.4 and 5.5. Respectively, the context information used for correction serve as an input in the preprocessing of the physiological signals. Context information used for improvement can be used in different steps of a model.

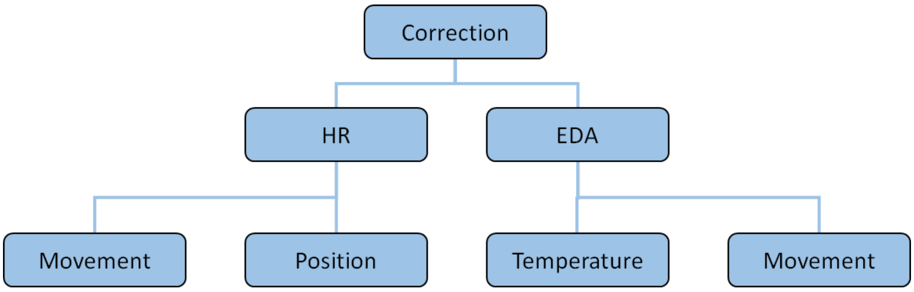


Figure 5.4.: Context information used for correction

Measures for correction of physiological data (shown in figure 5.4) fall into the class of controlling influencing factors of these physiological measures. For example, temperature and movement sensor integrated in the Q-sensor can be used to control interpretation of EDA values, which rise because of arm movement or changing temperature. For the heart rate, movement and position are important to control the interpretation. As introduced in chapter 2.3, heart rate can be influenced by position or movement of the body. In this case, the integrated step sensor of the smartphone is used to determine movement. The position, if an user is sitting or standing, might be determined with a combination of different sensors. The movement sensor integrated in the Q-Sensor collects information about arm movement. In combination with the step sensor sitting, standing and walking might be distinguished reliably to a certain level.

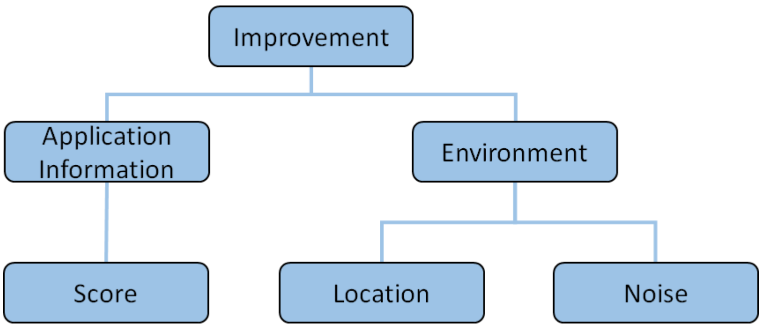


Figure 5.5.: Context information used for improvement of the interpretation of physiological signals

As mentioned, context information is also used for direct improvement of interpretation in the model. Context information, which falls into this category is shown in figure 5.5. Information, like e.g. location or movement in general, can be used to offer additional information to determine user state more accurate. For example, different user interfaces or information in the application can be offered based on location-awareness or difficulty of a game can be lowered based on performance. Context information of this category is not only received by integrated hardware sensors, but also by applications, like performance measures or a current score.

Some information can be used in both categories. For example movement falls into the category of correction because it can be used for correction of HR and EDA, when a person is moving. Movement falls also in the category of improvement because of the fact that movement can also be used for general improvement of current user state, e.g. adaptation of user interface, when a person is walking.

5.3.2. Location

Location can be determined in different ways, for example by GPS sensors, cell ID or Wi-Fi. The three mentioned ways for location determination have different advantages and disadvantages. GPS is very power consuming, depending on the refresh rate. Other differences between the sensors can be found in accuracy and time needed for refreshment of position. The accuracy of position estimation can depend on the different conditions. Bad weather or walls of a building might influence the accuracy of GPS. Within buildings, position accuracy of location acquisition over Wi-Fi might be better than GPS, depending on the situation in the building.

With help of GPS sensor, it can also be determined, if a person is travelling fast, e.g. by car or train. The technical details of implementation and integration into the model will be introduced in detail chapter 6 and 7.

5.3.3. Movement

An aspect regarding controlling the effects of the environment on the user is movement. By using built-in sensors like gyroscope or magnetometer, movement can be determined with help of different detection algorithms. Different operating systems, like e.g. Android, offer a specialized step application programming interface (API), delivering the steps taken without the need to implement detection algorithms.

If an user is moving, the distraction from using an application is bigger than when an application is used in a non-moving scenario. When moving, users have to split attention to pay attention to the environment to e.g. prevent accidents. The interpretation of affective and cognitive state can be enhanced, e.g. difficulty can be reduced, when walking.

When using physiological data, movement information is also needed to control influencing effects of movement on the data. For example, heart rate may rise when walking fast.

5.3.4. Context Information - Application Information

Depending on the application, several pieces of information might be useful to be integrated into the model. Performance statistics can be used in games or learning applications. In other applications performance measures like the time needed for a certain task may be an indication for current state. The value is not useful for control, but for improvement of user state interpretation.

Values for statistics have to be transmitted from the application to the interpretation model. This measure depends strongly on the type of application. Some applications may also not offer statistics or it might not be reasonable. The usage of statistics are further described in the section of the specific applications.

5.3.5. Other Context Information

Besides the previous described context information, modern smartphones offer sensors to measure even more context information. Examples for other sensors integrated in smartphones are proximity sensor, microphone, photometer or magnetometer. The proximity sensor allows to determine if a user is holding the phone near the face, the microphone measures the noise level and the photometer can measure the illuminance of the environment.

Based on these sensors, extensions of the model like identifying with help of the photometer if a user is sleeping (in a dark room) or not could be added. These information is not integrated in the model, as the model concentrated on movement, location and application information. But the possibility exists to integrate them.

5.3.6. Conclusion

As shown in this section, different context information can be collected. They can be used for control of measured physiological signals, as well as for improvement of the user state interpretation itself.

From the presented context information, location, movement and application information like performance are important for this work. Movement is primary used for correction of physiological signals, performance and location for improvement of user state interpretation.

For location, GPS sensor was selected for this work. As the power consumption might be higher than with the other sensors, a refresh of position information is only done occasionally. For determination of movement, the integrated step sensor API is used, having the advantage of not needing a separate implementation of step determination based on accelerometer etc. Out of the application information, performance was chosen to be a suitable measure for most scenarios. It might especially be a benefit for applications involving any kind of performance as e.g. games. The implementation and technical details will be described in chapter 7.

5.4. Output

As defined in the requirements analysis, to cover a broad range of different applications, affective and cognitive state are needed. In chapter 2.9 different psychological models were described for the determination of affective state. In the following, an overview of the concepts used for affective and cognitive state are given.

5.4.1. Affective Value

Different models have been used in affective and physiological computing to determine the affective state. As presented in the state of the art analysis in chapter 3, the number of different states, which have been distinguished, range between 2 and 8. In most cases, basic emotions or a model based on the two-dimensional valence-arousal approach, has been used.

In this work, it was decided to use a model with eight different affective states based on the two-dimensional valence-arousal model of Russell. The theory that valence and arousal span up a two-dimensional space is well established and widely used within the area of affective computing. The eight different states are based on the original circumplex model of Russell introduced in 2.7.1. As eight states might be too detailed for reliable interpretation of physiological signals when signals get lost or are corrupt, the eight states can be combined in a set of four states. This offers a more reliable interpretation, when needed depending on available signals.

The chosen sensors allow to measure arousal and valence based on EDA and HR. The literature showed, that Skin Conductance Level correlates to arousal (see 2.5.1). In the same way, HR is associated with valence. Besides HR, different context or application information can be additionally used for valence. The details of the interpretation and the eight different affective states will be presented in chapter 6.

5.4.2. Cognitive Value

In most studies, presented in the state of the art analysis, mental workload had been determined and distinguished into different levels. Mostly, mental workload has been measured based on a HRV analysis or different aspects of EEG measurement. EEG sensors are up to the date of this work not unobtrusive enough to meet the requirements for mobile scenarios. Based on the used heart rate sensors, a spectral analysis can be done to calculate mental workload. The results will be normalized and categorized into different level of workload: low, medium, high and very high. Categories have been chosen based on the Yerkes-Dodson curve, implying a higher performance under a medium amount of stress. Details of the analysis, the signal processing and the normalization of HRV will be described in the following chapter.

5.5. Method of Interpretation

As presented in the state-of-the-art analysis several models exist. All models have certain advantages and disadvantages. For the thesis, a model respecting as many aspects of the requirements as possible, has to be chosen. In the following, an overview of important aspects and the grade to which the different models meet them, is given. Finally, a model for interpretation is chosen.

5.5.1. Learning and Configuration

Learning and configuration are an important point. As described in the requirement section, learning and configuration needs to be fast and flexible. Neural networks and support vector machine (SVM)s are able to learn the classification of data with help of different algorithms and methods. The learning process requires data sets for training. As presented in chapter 3.2.1, different kinds of neural networks, the recurrent networks, are furthermore able to reevaluate and adapt during runtime.

In comparison to SVMs and artificial neural network (ANN), fuzzy logic itself has not a learning process. The rules are defined by an expert. As introduced in chapter 2.9.3, this rules can be expressed with "IF... THEN.." rules in a linguistic way.

As Mandryk and Atkins [MA07] state, an advantage of fuzzy logic models is, that no data of all possible states is needed in comparison to ANN. Different affective states are hard to measure in studies. An example is an extreme affective states like sadness.

5.5.2. Classification Rate

As stated in the conclusion in chapter 3.2.5, SVMs achieve the best classification rate. The studies showed, that the classification rate of SVM and ANN improved, when the number of training data sets increased.

Fuzzy logic on the other hand was evaluated in studies by determination of least squared error. In the study of Rani et al. [RSSA03] the least squared error was relative stable and only differ slightly between different number of data sets. The other evaluated approach based on a regression tree, was slightly better in classification rate under good conditions, but with only few data sets, fuzzy set approach outperformed the regression tree approach.

5.5.3. Flexibility and Robustness

During usage in mobile scenarios, input channels might change. Additional ones could get connected or connection to established channels can get lost. Neural networks would have to change the weighting of nodes, implying a new learning step of new data sets. The same would apply for SVMs. Fuzzy logic systems on the other hand can be configured in advance by an expert, offering different rule sets to cover different sensor configurations.

When adding completely new input channels the models need to be modified. When using Fuzzy Logic, new rules have to be defined for the additional input channels by an expert. The other approaches need training data sets to learn the new configuration.

A further advantage of fuzzy logic is the handling of noise. As Novak et al. state [NMM12] the inter- and intrasubject variability can create noise, which can be handled by fuzzy logic approaches in comparison to the other introduced methods, leading to better and stable results.

5.5.4. Conclusion

Analyzing the different aspects, the models have different advantages and disadvantages. None of the introduced models meet all criteria. For this work, a fuzzy logic approach has been selected. The selection was made based on the advantages of fuzzy logic approaches in configuration and handling of noise, as these points are important aspects in mobile scenarios. Classification rate of SVMs might be better, but the values achieved with fuzzy logic approaches still seem to be sufficient and may be increased in combination with context information.

The applied method, its functions and variables, will be described in detail in the following chapter 6, the implementation of the concept in chapter 7.

5.6. Applications or Mobile Scenarios

For proof of concept, different mobile applications are needed. To test the affective state as well as the cognitive state, applications for entertainment are needed, as well as applications managing and using the cognitive state for adaptation. Examples for entertainment are games, which have been widely studied with physiological input, but mostly not in mobile scenarios. For cognitive state adaptation, applications used under a high amount of stress or for learning, are used. Many state of the art work in the area of adaptation based on cognitive state is from the area of workload management in flight and aeronautics, which may not be ideal candidates for mobile scenarios. Adaptation of the current cellphone state, like presented in the interruption management example in chapter 3.3.3, seems a more promising approach for cognitive state in mobile scenarios.

The applications get the information of affective and cognitive state by the model. The decision about how and when something in the application will be adapted is done in the application itself to allow a high degree of flexibility. For this decision, simple rule-based systems can be integrated in the applications. For that reason, the model needs to communicate the results of the user state interpretation in an understandable format.

Developed applications and their mechanism for adaptation will be described in detail in chapter 8. First studies and evaluations with these applications will be described in chapter 9.

5.7. Concept

The in chapter 3.1.1 introduced biocybernetic loop, is the heart of many adaptive applications in the area of physiological computing. In this chapter, the process of the loop for the scope of this work, will be presented, as well as a detailed model based on the decisions made in this chapter. The model is divided into several steps, reflecting the different stages of a biocybernetic loop.

5.7.1. Biocybernetic Loop

The process of the biocybernetic loop can be divided into two main parts besides the user in this work: the application itself and the user state classification. As shown in figure 5.6, the biocybernetic loop involves in the first two steps, the signal processing of the physiological and context signals as well as the interpretation and classification of the user state, based on the input.

On the application side, a feedback controller handles the incoming values of affective and cognitive state and chooses an adaptation. Afterwards, the application realizes the adaptation, which is presented to the user. The concept is flexible enough to realize negative, positive or mixed feedback loops.

The decision, if a loop is positive, negative or switches is controlled within the application. The user state classification part only processes and interprets the data.

5.7.2. Overview

To meet and address the different points collected in the requirements analysis for mobile scenarios, a model has been developed for structuring the different parts of the solution into different steps. The model, shown in figure 5.7, consists of three steps, separating the

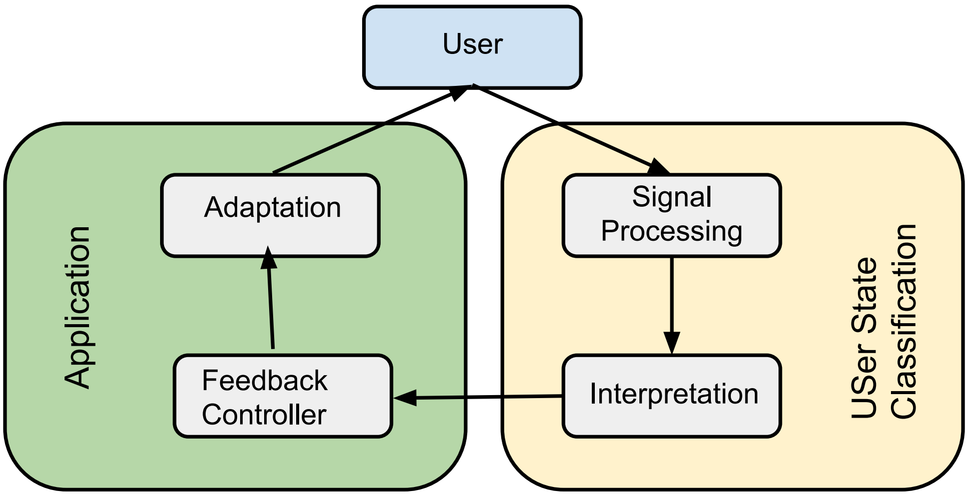


Figure 5.6.: Biocybernetic loop for the concept

solution in "Signal Preprocessing", "User State Interpretation", and "Application". These separation of the problem in different steps allows a higher flexibility if one part of the model is changed or removed.

The preprocessing step receives the different input signals from the external sensors and the smartphone itself. The signals are then checked for corruptness and availability. Several steps, like spectral analysis or normalization are done, before the transmission of the values to the user state interpretation.

The received values in the user state interpretation get transformed with two fuzzy logic systems. In the first fuzzy logic system, valence, arousal, mental load and context are determined. The second fuzzy logic system finally transforms them to cognitive and affective state.

These values are transmitted to the application. The application decides then about adaptation and represents the result to the user. Besides the cognitive and affective value, the processed physiological signals from the signal processing step are also available for the application. This allows to use physiological signals directly for direct and indirect control of applications.

The most important steps, models and applications of the single components of the different steps will be briefly explained in following subsections of the this concept chapter and in detail in chapter 6.

Data Channels

As input are several data channels used. In this work physiological signals and several in smartphone integrated sensors are used. Sensors for measurement in mobile scenar-

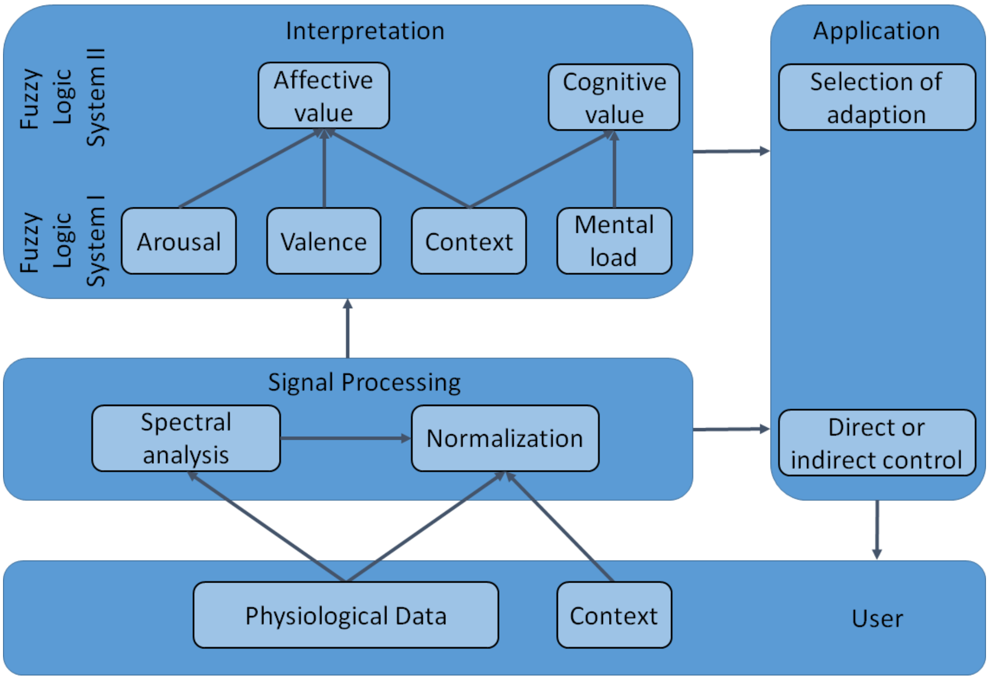


Figure 5.7.: Architecture of the model consisting of four different layers

ios have to fulfill several requirements, which will be taken into account when choosing physiological signals and available sensors. The requirements, the choice of physiological signals and the chosen sensors are in detail introduced in chapter 5.2.

Signal Processing

In the signal processing step, different algorithms for spectral analysis and signal preprocessing need to be implemented. After preprocessing and power spectrum density calculation, the results of power spectrum density (PSD) and EDA get normalized.

Besides transmitting the processed values to the interpretation part, the signals are also transmitted directly to the application, if the application wants to use the input directly, for example for direct control of applications by physiological signals. The single steps of the signal processing will be presented in chapter 6.2.

Affective Value

Affective value is calculated based on physiological signals and context information. It reflects an abstract value of the current emotional state of the user. This value is especially useful in applications like games, where the affective aspect is in the foreground of the application. The processing of the input channels, calculation of the output and the different possible results will be further discussed in chapter 6.4.

Applications can adapt, based on the output of Affective Value or can ignore this value and choose to use one of the other values for adaptation based on the usage scenario of the application.

Cognitive Value

The cognitive value reflects the current cognitive state of a user, based on mental effort and context information. Especially in applications like e.g. in learning or other scenarios where performance is the most important aspect, cognitive value is useful.

Like the affective value, this value might not be suitable for every application depending on the application scenario. Developers of certain applications have to decide if they use or discard this value.

Context Information

Context information is collected by different sources. On the one hand input from different sensors, integrated in the smartphone. On the other hand information from the application itself, like performance measures or usage statistics. This information can be used to help processing affective and cognitive state in the same layer or as a direct input into the application input logic layer.

A further description of the different context information types and how they are used will be given in section 6.5 and 6.4.

Application

The application consists of two different parts in the concept. On the one hand a feedback controller, that chooses adaptation based on affective or cognitive state. On the other hand, a connection to the signal processing unit is established. With help of this direct input, applications that adapt to the direct physiological parameters can be realized. Examples here fore are e.g. visualization in games, which synchronize with heart rate or gaming modes that allow a direct influence of the game when using heart rate.

6. Model and Input Channels

In this chapter, the components of the concept will be described further. In a first step, an overview of the user interpretation part of the model will be given. The model used for transformation of physiological data to affective and cognitive state will be described. Preprocessing of signals and the single steps of this process for transforming physiological data in to affective and cognitive factors like valence, arousal and mental effort will be introduced. Context information will be categorized and the usage of these in the model will be explained. These values are transformed to an affective and a cognitive state of the user. Both steps are based on fuzzy logics which will be described as well. Finally, the parts of the model, handling robustness and reliability are described and a conclusion is drawn.

6.1. Model Overview

Continuing the model overview from chapter 5, more details of the interpretation of physiological and context information is given in this section. The model shown in figure 6.1 is based on the fuzzy physiological approach of [MA07]. As described in chapter 3, the model is suitable for continuous modelling of physiological data and has different advantages for mobile scenarios (see 5.5 for further details).

In comparison to the work of Mandryk and Atkins [MA07] the model differs in some aspects. The input channels differed, as in mobile scenarios not every measure is suitable for mobility due to available wireless sensors. Instead of heart rate (HR), Electrodermal activity (EDA) and electromyogram (EMG) the signals heart rate (HR), Heart Rate Variability (HRV) and Electrodermal activity (EDA) were used as physiological input. Additionally context information was added.

The original model concentrated on games, where the aspects boredom, challenge, excitement, frustration and fun were chosen as output. In the model presented in this work, more aspects are included as the range of applications is broader. Therefore input is not only transformed into valence-arousal space but also mental effort was calculated for a cognitive state of the user. In a second step, valence, arousal, cognitive aspects and context are transformed into affective states and a cognitive state.

In the first step, signals are fuzzified by membership functions (see 6.3 for details). Afterwards fuzzy rules transform these value to a membership value in valence, arousal and mental effort (6.4.1 and 6.5.1). These value are then combined with context information and used as an input for the second step. In the second step, the values are transformed to values for affective and cognitive value with help of fuzzy rules (6.4.2 and 6.5.2).

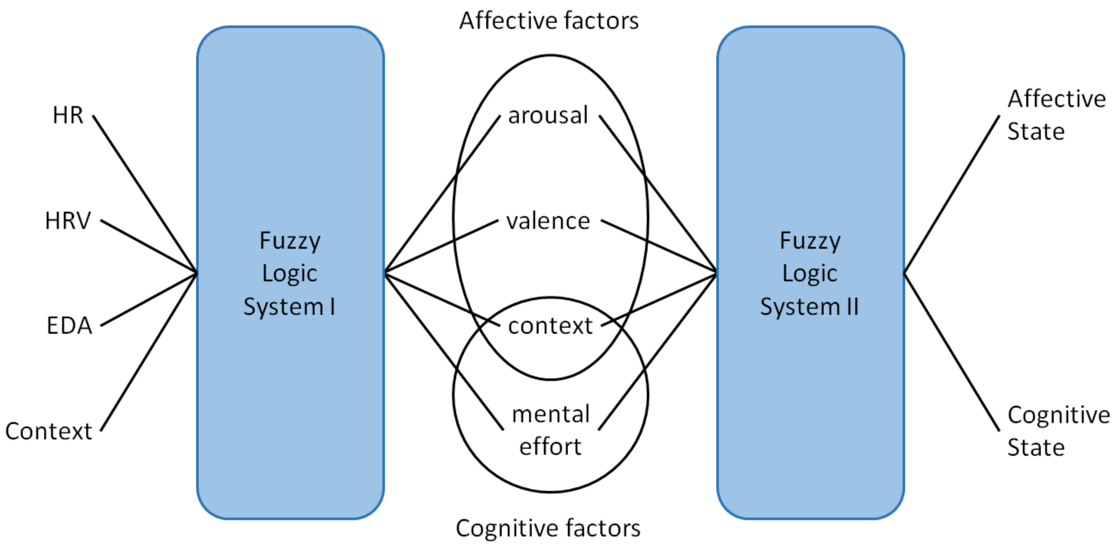


Figure 6.1.: Overview of the processing of input by fuzzy rule sets to the final output for affective and cognitive state

In the first step of data processing, context information is used to correct interpretation of physiological data. In the second step it is used to improve the determination of different states. In the following sections, the single steps of the model will be explained in detail. For a better understanding of the overall process of affective and cognitive state, each will be described in the whole with the different steps instead of separating the steps of the model. The following section will start with the preprocessing of the input, followed by the processing of context information. After that, the parts of the model for affective state will be described followed by a section for cognitive state.

6.2. Preprocessing of Physiological Signals

After receiving the physiological signals, several steps have to be taken for preprocessing of the signals. The signals have to be checked for corruptness and depending on the measure to be normalized for further processing. In the following, the necessary steps for EDA, HR, HRV and context information are presented.

6.2.1. Electrodermal Activity

In a first step, a baseline has to be determined for normalization. Baseline can be measured during a few minutes of resting time. For normalization, maximum value is also needed. To determine maximum value, Lykken and Venables [LV71] suggest to let participants blow up a balloon until it bursts. Another possibility, if it is not possible to measure maximum values, is to take the current maximum value and refine it as soon as higher values have been measured. When using this approach, accuracy might not be high at the beginning, but improves over time. As introduced in chapter 2.2.3, skin conductance level is normalized with the following formula:

$$SCL_{norm}(t) = \frac{SCL(t) - SCL_{min}}{SCL_{max} - SCL_{min}} * 100 \tag{6.1}$$

Corruptness of the signal is checked by two aspects: battery lifetime and by checking the signal availability itself. Depending on the sensor, false data may be transmitted when battery gets low. To prevent this case, battery lifetime of Q-Sensor is regularly checked, as the values of battery life time are transmitted by the application programming interface (API).

6.2.2. Heart Rate

To normalize heart rate, the age of the user is needed. As described in chapter 2.3.4, Tanaka, Monahan & Seals [TMS01] developed a formula for calculating maximum heart rate:

$$HR_{max} = 208 - (Age * 0,7) \tag{6.2}$$

As the heart rate may never reach the maximum value, except in the case of users doing sports, the maximum was set at 80% of HRmax as at this level the anaerobic zone starts [BBC16]. To determine the minimum value a baseline is measured during a resting period in a sitting position. In the same way, as the SCL is normalized, heart rate is normalized with:

$$HR_{norm}(t) = \frac{HR(t) - HR_{min}}{(HR_{max} * 0.8) - HR_{min}} * 100 \tag{6.3}$$

The signal is checked for corruptness by checking if the different successive heart beats are plausible. If they have to high variation, signal is very likely corrupted.

6.2.3. Heart Rate Variability

The HRV is given in milliseconds between two R-Peaks of heart beats. As described in chapter 2.3.5, mental effort can be determined by a spectral analysis of HRV intervals. The power of low frequency in the band of 0.06 - 0.14 Hz reflects the autonomic system and mental load. The result of the integration of power spectrum in this band is normalized by a division with baseline measurement results. When subtracting the result from 1 the result is a value between 0 and 1, where 0 represents a low mental load and 1 a high mental load. This can be expressed by the following formula:

$$mentalLoad(t) = 1 - \frac{PSD(t)}{PSD_{baseline}} \tag{6.4}$$

As the signal has already been checked for corruptness in the heart rate preprocessing, the signal does not need to be checked in this step.

6.2.4. Context Information

Preprocessing of context information includes the normalizing of steps for movement and identifying the location with help of predefined areas. Performance information from the application itself already have to be normalized, transmitting a value between 0 and 1, where 0 relates to a low and 1 to a high performance.

Steps are normalized to a value between 0 and 1. As maximum value for steps are 140 steps per minute used, which correlates to jogging. Everything beyond is not of interest, as it might not be reached in usage scenarios. Normal walking has around 70 to 75 steps a minute. Steps are normalized by dividing the current count of steps by the maximum of 140 steps a minute:

$$steps_{normalized}(t) = \frac{steps(t)}{140} \tag{6.5}$$

6.3. Fuzzyfication of Input

Context information is used for both, affective and cognitive state. To be integrated in the model, movement gets fuzzified as well as performance, if available.

6.3.1. Movement

In the preprocessing, steps were normalized to a movement value between 0 and 1. Based on this value, movement is fuzzified into three sets: low, medium and high. The membership functions are shown in figure 6.2.

Low movement corresponds to no or only slow movement, where as medium corresponds to activity like walking and high to fast walking or running. The membership functions have been defined based on tests of different conditions of movement and statistics. The area for low movement has been defined relatively small, as different parameters rise relatively fast with even low movement.

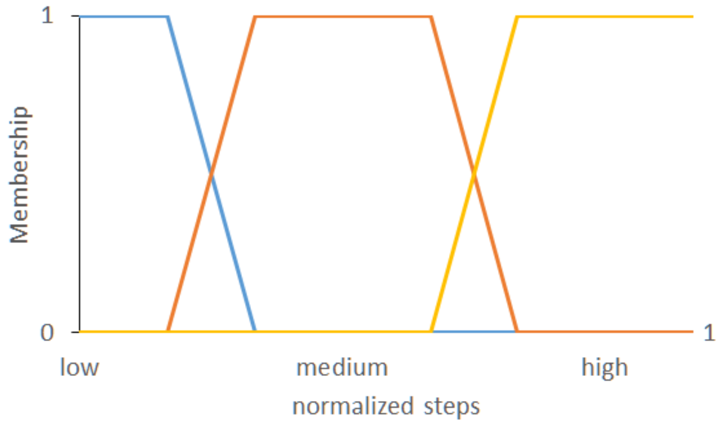


Figure 6.2.: Membership functions for modelling the movement

The movement variable is used in the first step of determine affective and cognitive factors as well as in the second step to determine the output.

6.3.2. Performance of Application Usage

A normalized performance parameter can be transmitted by an application as described in the preprocessing step. The normalized value is then transformed by a membership function (shown in figure 6.3) in one of five different states: very low, low, medium, high and very high.

The five classes are equidistant, as the performance is already normalized. Performance is used for valence in the first step and in the second step for affective and cognitive state.

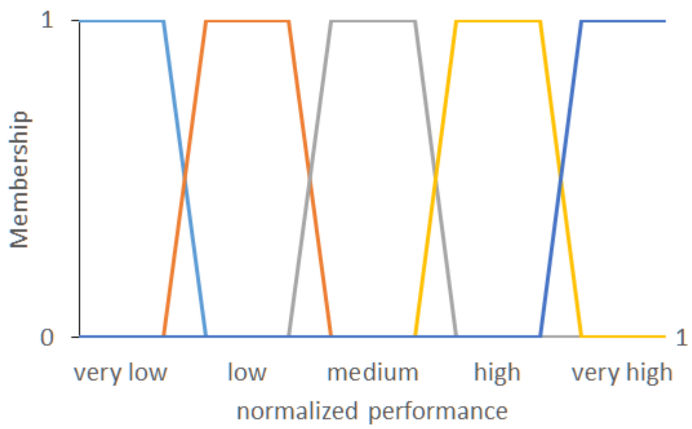


Figure 6.3.: Membership functions for modelling the performance

6.3.3. Physiological Signals

After the preprocessing step described in 6.2, the values are normalized. For fuzzification of the signals, membership functions have been created. The membership function for EDA is shown in figure 6.4.

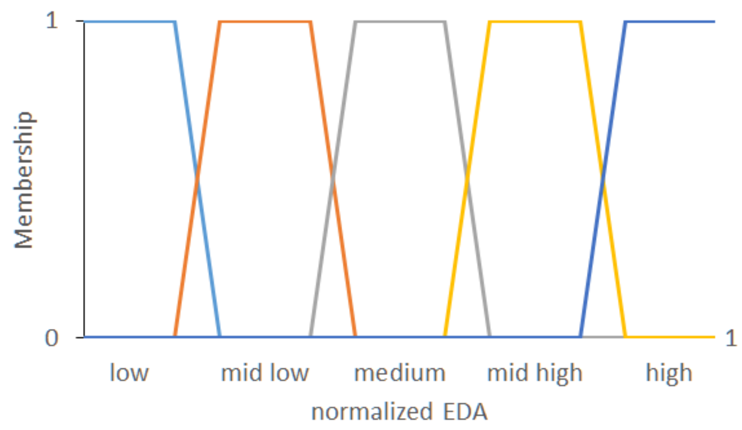


Figure 6.4.: Membership function for fuzzification of electrodermal activity

The five classes are equidistant based on the normalized value. Five classes have been chosen to cover a certain width of states. A finer granularity would not lead to better results, as this would require a more accurate and controlled measurement of the signals.

The membership functions for EDA and HR are shown in figure 6.5.

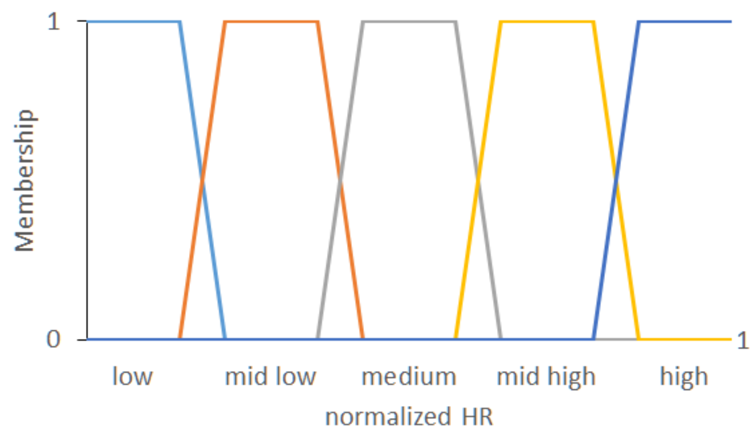


Figure 6.5.: Membership function for fuzzification of heart rate

As described in the preprocessing section, power spectrum density (PSD) of low frequency area is normalized after spectral analysis. The normalized value is in a first step transformed to a fuzzy variable. Same as heart rate and electrodermal activity, HRV is categorized into five classes in figure 6.6.

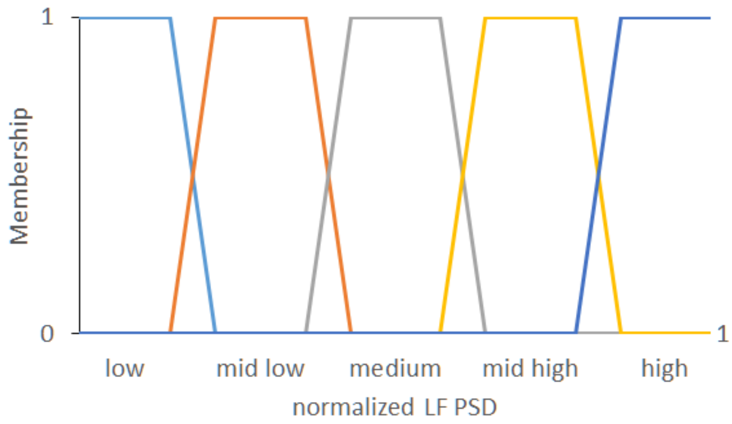


Figure 6.6.: Membership function for fuzzification of LF PSD results

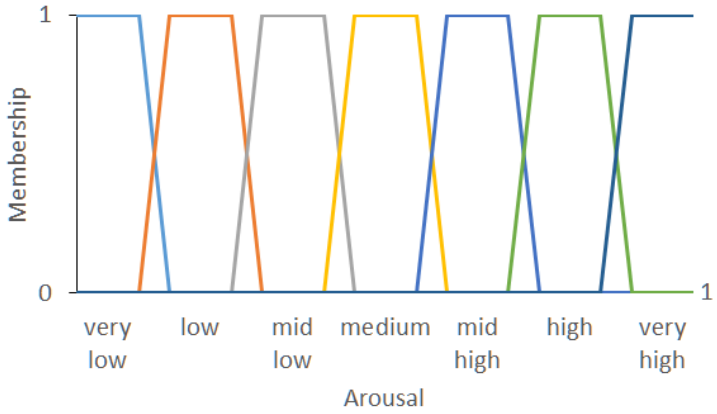
6.4. Affective Value

The affective value is based on EDA, HR and context information. Due to the fact that the interpretation is based on the valence-arousal model, the physiological signals and context information get transformed in a valence and an arousal value. These values are transformed in percentages of membership in different affective states in the second step.

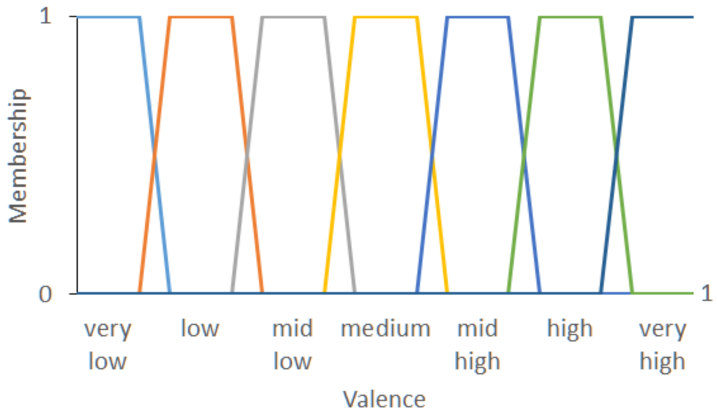
6.4.1. Transformation to Valence and Arousal

Valence and arousal are quantified with help of a system of fuzzy rules based on the results of HR, EDA and movement. As the base for determination of affective state will be a 7x7 affective grid, which will be described in detail in the following subsection, membership functions of valence and arousal (see figure 6.7) are each divided into seven classes: very low, low, mid low, medium, mid high, high and very high.

Context information was used to control influencing effects on HR and EDA. To control influencing factors, movement was used. To improve the interpretation of valence, performance is integrated in the fuzzy rules. The fuzzified values of movement and performance were integrated in the fuzzy rules. In total 15 rules for valence and 10 rules for arousal were created. They were based on the psychophysiological background introduced in chapter 2. HR correlates to valence, as well as EDA correlates to arousal. The rules for arousal are shown in the following table 6.1:



(a) Arousal



(b) Valence

Figure 6.7.: Membership functions of valence and arousal

if (EDA is high) then (arousal is very high)
if (EDA is high and movement is not low) then (arousal is high)
if (EDA is mid high) then (arousal is high)
if (EDA is mid high and movement is not low) then (arousal is mid high)
if (EDA is medium) then (arousal is mid high)
if (EDA is medium and movement is not low) then (arousal is medium)
if (EDA is mid low) then (arousal is medium)
if (EDA is mid low and movement is not low) then (arousal is mid low)
if (EDA is low) then (arousal is mid low)
if (EDA is low and movement is not low) then (arousal is very low)

Table 6.1.: Fuzzy rule set for interpreting arousal based on EDA and movement

The results are the memberships in the different arousal and valence sets. The whole set of fuzzy rules for interpretation of valence can be found in the Appendix.

6.4.2. Processing Affective State

Valence and arousal reflect the affective state of an user and are used as input in the second step together with context information. The output of this step are eight different affective states. The transformation and rule sets are described in the following.

The processing of the affective state is based on the circumplex model of Russell and the affect grid, described in 2.7.1 and 2.7.2. In the original model of Russell [RWM89], eight different words, describing an emotional state, are plotted as points on a circumplex graph. The original affect grid is a 9x9 grid, with valence and arousal axis. In this work, the affect grid was reduced to a 7x7 grid for simplification. This 7x7 grid was divided into eight different areas, by projecting the eight states of the Russell circumplex on the grid. Every area is described with 2 words with help of the extended Russell circumplex model, which had 32 different emotions on the circumplex. The result is shown in figure 6.8.

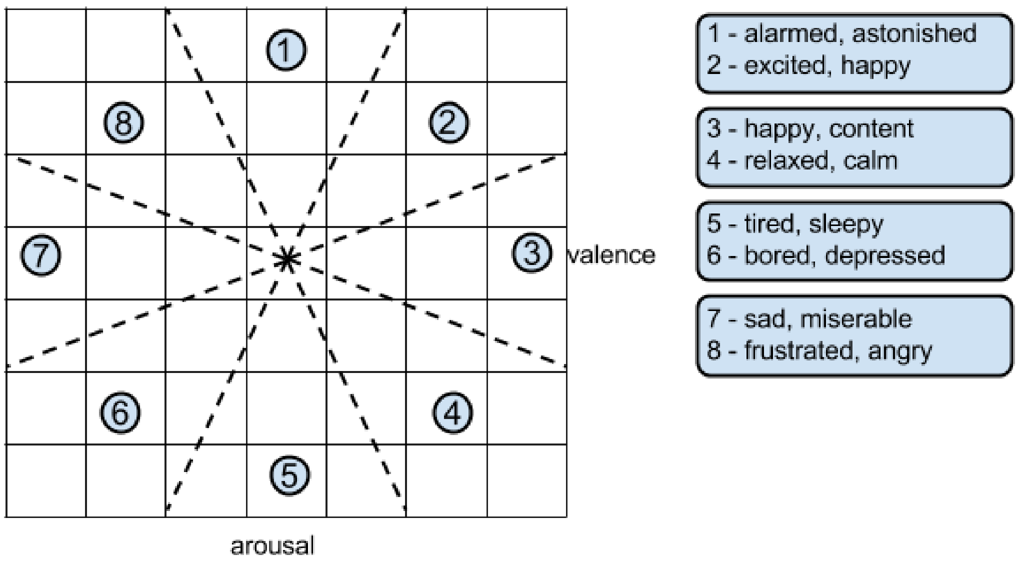


Figure 6.8.: Model for the interpretation of valence and arousal to eight affective states

Eight states allow a detailed reflection of the affective state in mobile scenarios. A further possibility is the reduction to 4 states, clustering two states to one state. This could be done like shown on the right side of figure 6.8. State 1 and 2 can be clustered to a positive and active state, as well as state 3 and 4 can be clustered to a positive but inactive state. The same is possible for state 5 and 6, a negative and inactive state, as well as 7 and 8, a negative and active state. This simplification to four states offers a solution in robustness, when not enough signals are available to distinguish between 8 states to offer enough reliability

Valence and arousal were transformed by 146 fuzzy rules to the eight affective states (see Appendix B.4). The membership functions were generated based on matching physiological data and subjective rating of previous studies. Like in [MA07] the output of the membership functions generates four different levels: very low, low, medium and high. The membership function of state 1 (alarmed, astonished) is shown in figure 6.9. Very low has only a small portion of possible membership, as it reflects the areas where a certain affective state is nearly not occurs.

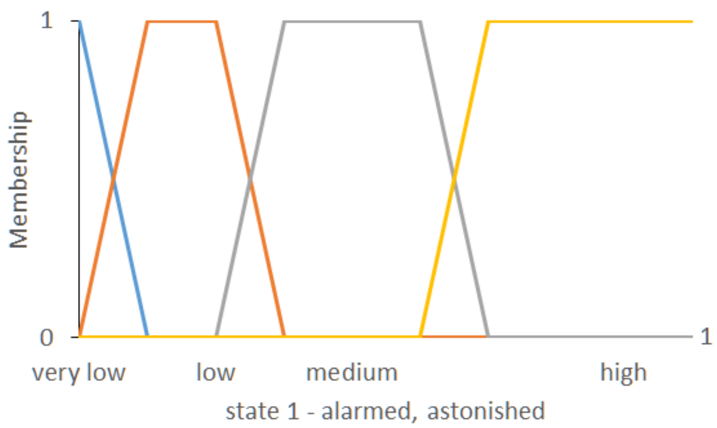


Figure 6.9.: Membership function for modelling affective state 1 (alarmed, astonished)

As the eight different affective states are very similar in the following processing steps, the steps will be shown only for state 1 (alarmed, excited) as an example. The membership functions as shown in figure 6.9 also apply for the other affective states.

Besides the membership function, the affective state can also be presented in the affective grid. Figure 6.10 shows the representation of affective state 1 in the affective grid. High membership is represented in dark blue, which are cells of the grid that are completely within the area of the affective state. Medium membership is represented by blue, which are cells that are at least 50% and less than 90% inside the affective state. Low membership is represented in light blue, which are cells that are less than 50% but at least 10% inside the affective state area. Cells with less than 10% have a membership of very low.

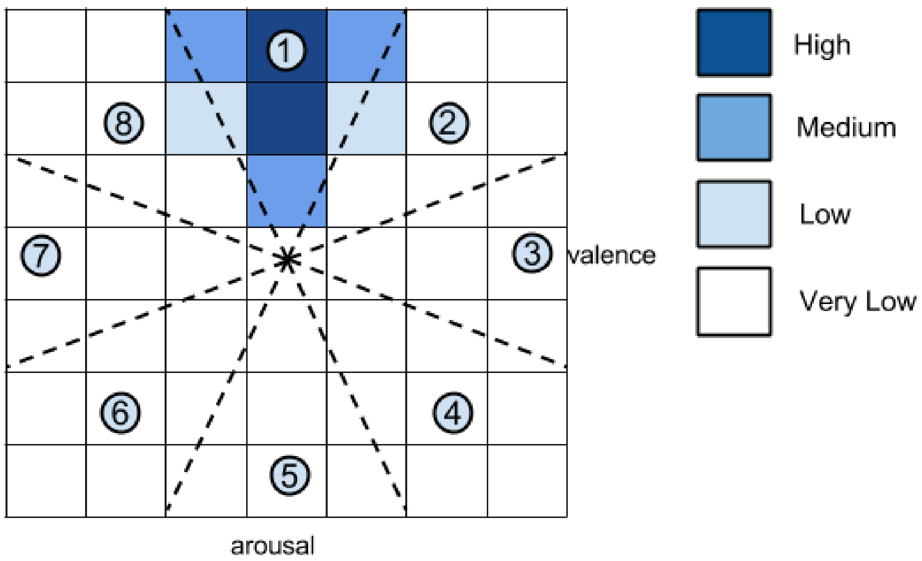


Figure 6.10.: Representation of affective state 1 in the affective grid, with the different memberships

After calculation of the memberships in the different affective states, the output needs to be defuzzified to have a crisp value for the applications. Different approaches for defuzzification exist. A widely distributed method is the center of gravity approach [Ros09], which will be used in this work.

6.5. Cognitive Value

The cognitive value is mainly based on HRV and reflects how cognitive occupied an user is. HRV and its relation to the psychological counterparts are in detail described in chapter 2.3.5 and 2.4.1. The two processing steps to get the cognitive state of the user are described in the following subsections in detail.

6.5.1. Processing Mental Load

In the second step, cognitive value is processed based on the fuzzified PSD values of the first step. Even so, no other signals are influencing the mental load value and mental load could have been directly transformed in this case, the step is done to offer the possibility of adding new data channels for mental load to the model in the future, e.g. electroencephalography (EEG) measures.

Membership functions transforms power spectrum density and context information to mental effort, which is shown in figure 6.11. The output of mental load is categorized in four different states: low, medium, high and very high. A set of fuzzy rules transforms the fuzzified results of PSD to a mental load value.

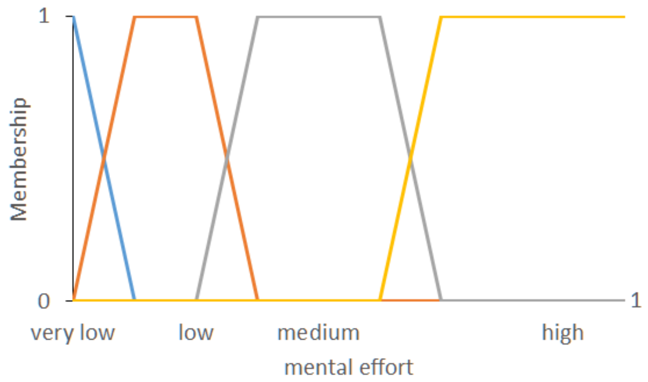


Figure 6.11.: Membership function for transformation of LF PSD results to mental load

The shape of the membership function is based on the curve of Yerkes and Dodson, presented in chapter 2.7.4. Based on this curve, the best performance would be reached in a medium stress load, in this case mental stress. In low and high areas, performance decreases. The high area was divided in two parts, to offer more details on high mental load level, as they are typically more interesting for most applications in comparison to the low mental load area.

In total five rules were used for transformation. The full set of rules is shown in appendix B.3. The rules were based on power spectrum density solely.

6.5.2. Calculating Cognitive State

In the end of the processing, the cognitive value represents a combination of mental effort based on HRV, environment and performance statistics. This value might not be important for every application, but is important for applications where a user has to perform well in a given task for work purposes instead of fun. The membership function for cognitive state is shown in figure 6.12.

The membership functions are identical to the membership functions for mental load in the previous section. Cognitive state is divided into four states: low, medium, high and very high. Like outlined in the mental load processing section, a further distinction of low cognitive states might not be as interesting as high cognitive states.

Context information which is used are performance for statistics of success or fail in an application (e.g. learning applications) and movement to improve the interpretation of power spectrum density further or to give a tendency in which direction cognitive state might go. When moving, a user may be distracted by the environment, lowering the cog-

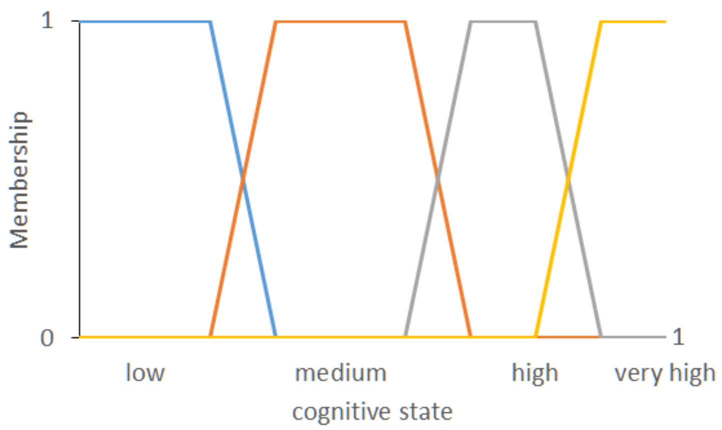


Figure 6.12.: Membership function for transformation to cognitive state

nitive availability. In this case, mental load might be higher as it would normally be in a not-moving situation. Performance is used to identify the cognitive state in different situations. For example, when mental load is high and performance very low, cognitive state is very high.

In total, 15 fuzzy rules have been created, shown in table 6.2. The rules are based on principles and observations made in studies, presented in chapter 2.

6.6. Handling Robustness

As one of the theses focuses on robustness, several steps have been taken to ensure robustness of different aspects of the model. Key points for robustness are especially loss of a signal, corrupted signals and influencing factors of the physiological signals.

Detection of corruptness and signal loss are handled in the preprocessing step. The details for detection are described in chapter 6.2. To handle this case, two approaches are integrated. In a first step, it is checked, if lost or corrupted signals can be replaced by other input channels.

In the second approach, if one important input channel is lost or corrupted and cannot be replaced, the output states can be reduced to allow more reliable results. As described in section 6.4.2, the eight affective states can be reduced to four states, allowing more reliable results for these four states, instead of less reliable results for eight different states

Besides signal loss or corruptness, influencing factors on physiological signals have to be controlled to improve robustness and reliability of the model. Especially heart rate is easily influenced by movement. Based on context information, like e.g. step sensor, the physiological signals can be controlled, which is integrated in the preprocessing step.

6.7. Conclusion

In this and the previous chapter, the model and its processes were presented in detail. The single steps, from input to interpretation of user state and the intermediate steps. The model starts with a preprocessing of the signals combined with first steps for improving robustness. After this, signals were transferred in two steps to an affective and a cognitive state with a fuzzy logic approach.

The calculation of affective state is based on the affect grid and the circumplex model of Russell. Cognitive state is mainly based on a spectral analysis of HRV combined with context information. In the first step of the fuzzy logic approach, input signals were fuzzified and afterwards transformed with fuzzy rules to values for affect, valence and mental load. These values together with context information were then again transferred with fuzzy rules to an affective and cognitive state.

The model addresses the defined requirements (see 5.1) of mobile scenarios and offers several approaches for improvement of robustness under different situations. In the following chapter, the implementation of the proposed model will be presented, followed by a chapter introducing several different applications in which the implementation of this model was integrated.

if (mentalLoad is very high) then (cognitiveState is very high)
if (mentalLoad is very high and performance is very high) then (cognitiveState is high)
if (mentalLoad is very high and performance is high) then (cognitiveState is high)
if (mentalLoad is high) then (cognitiveState is high)
if (mentalLoad is high and movement is high) then (cognitiveState is very high)
if (mentalLoad is high and performance is very low and movement is low) then (cognitiveState is very high)
if (mentalLoad is high and performance is very low and movement is not low) then (cognitiveState is very high)
if (mentalLoad is medium) then (cognitiveState is medium)
if (mentalLoad is medium and movement is high) then (cognitiveState is high)
if (mentalLoad is medium and performance is very low and movement is low) then (cognitiveState is low)
if (mentalLoad is medium and performance is very low and movement is high) then (cognitiveState is high)
if (mentalLoad is low) then (cognitiveState is low)
if (mentalLoad is low and movement is high) then (cognitiveState is medium)
if (mentalLoad is low and performance is high) then (cognitiveState is medium)
if (mentalLoad is low and performance is very high) then (cognitiveState is medium)

Table 6.2.: Fuzzy rule set for interpreting cognitive state based on mental load and performance

7. Implementation

Chapter 5 and 6 described the architecture, model and details of processing physiological input and context information into different user states. This chapter describes the implementation and realization of these concepts from first steps like connecting sensors to the final output and its forwarding to the applications itself. The requirements that are needed to be fulfilled by the applications to receive and process the information are also introduced. The implemented engine will be called MUSE, standing for ”mobile user state estimation”.

7.1. Development Environment

When developing for mobile devices, many different software development kits (SDK) and operating systems exist. The most popular ones are Android, Apple iOS, Windows Phone and Blackberry OS. Developed native applications are not compatible across the operating systems. The only way to run an application on all of these platforms would be a web-based solution.

Several aspects and requirements lead to the decision of an implementation for mobile devices running Android operating systems. A web-based solution was not a possible solution, due to the integration of external and internal sensors. At time of development, Android SDK had several advantages compared with other operating systems like Windows Phone or iOS. One advantage was the support of Bluetooth in the SDK, which is needed by the chosen sensors and the wide distribution of Android OS. In August 2014 Android OS had a market share of 84.8% [IDC16]. Bluetooth was early integrated in the SDK of Android. Apple’s iOS supports Bluetooth LE since version 6, which was released in 2012.

For testing and debugging, a LG Nexus 5 has been used. The Nexus 5 was one of the first Android Smartphones supporting the Bluetooth Smart standard and the new step sensor application programming interface (API) available since Android 4.4.

7.2. Architecture of the MUSE Engine

As proposed in chapter 5, the processing and interpretation of physiological data and context information is done in a background service due to limited resources on smartphones. Background services have the advantages of running the whole time in the background, independent of the applications that use them.

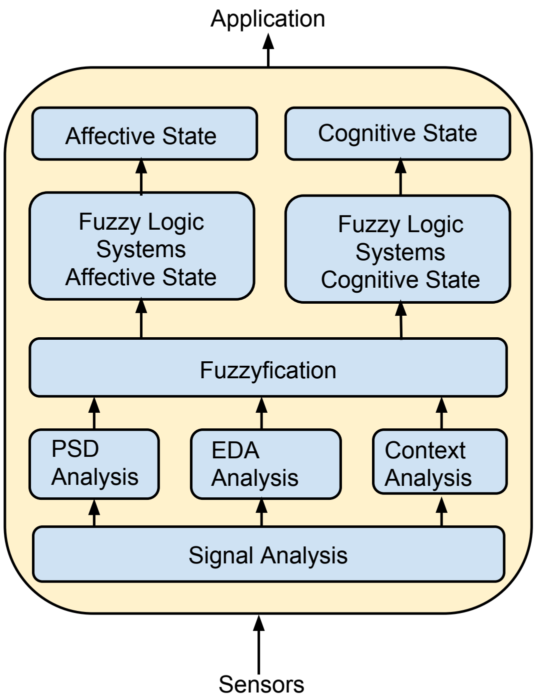


Figure 7.1.: Architecture of the background service

Within the background service, some steps are running simultaneously during the execution of the service. When the service is running, processing of data and calculations are constantly done. The most important functions, which are executed simultaneously are power spectrum density (PSD), Electrodermal activity (EDA) and context analysis. PSD Analysis is done, as soon as enough data is collected. EDA and context analysis is done from the start. The architecture of the background service is shown in figure 7.1.

Integration of sensors will be described in section 7.3, followed by a presentation of the data processing steps of physiological signals and context information in section 7.4 and 7.5.

7.3. Integration of Physiological Sensors

The sensors that have been chosen for measurement of EDA and heart rate (HR) have been introduced in chapter 5.2. The sensors are connected over Bluetooth. Between the sensors is a difference regarding the kind of Bluetooth connection. Q-Sensor and Zephyr HxM work with the Bluetooth 2.0 standard, while Polar H6 supports the newer Bluetooth 4 with Low Energy. Implementation of these two connections differ. Both are further described in the following in the sections of the respective sensor.

7.3.1. Zephyr HxM

The Zephyr HxM comes with a SDK and a documentation of the format in which the data is transmitted. As mentioned before, Zephyr HxM works with Bluetooth 2.0 standard. Initialization of connection is very similar to the Q-Sensor.

The HxM transmits data packets with several information every second. The sensor can only transmit data but not receive it. Each packet consists of the heart rate, speed, distance and an array with the last 15 heart rate time stamps. With these time stamps, RR intervals can be calculated.

7.3.2. Q-Sensor

Affectiva Q-Sensor transmits the measured data via a Bluetooth serial port for communication. Data is transmitted in a packet format as a stream of a string. Within this string, information from the integrated sensors is transmitted. The information of the motion sensor is encoded in a z, y and x value. Furthermore battery lifetime and temperature besides EDA, which is given in microsiemens. The structure of the packet looks like shown in figure 7.2.

*0-9**	*Gs*			*Volts*	*°C*	*uS*
Packet #	**Z**	**Y**	**X**	**Battery**	**Temp**	**EDA**
6	0.40	0.32	0.82	3.55	32.9	0.068

Figure 7.2.: Q-Sensor packet format [Aff11]

This packet format is transmitted in form of a stream of a string, which looks like the following:

0,0.39,0.12,0.86,3.55,32.9,0.068 r n
1,0.40,0.07,0.90,3.55,32.9,0.071 r n

2,0.39,0.23,0.95,3.56,32.9,0.071 r n
3,0.39,0.21,0.86,3.55,32.9,0.071 r n

Once the sensor is paired with the device, it gets connected with a Bluetooth Socket. In the next step an RFComm Socket constantly receives the stream. The string is parsed and information relevant information forwarded to the processing of the signals.

7.3.3. Polar H6

The Polar H6 supports the Bluetooth 4 standard with low energy. The Bluetooth 4.0 standard offers several generic attribute profiles (GATT). One of this GATT profiles is a specialized profile for heart rate monitors [SIG11], which is supported by the Polar H6. Based on the heart rate profile, different information can be accessed.

Due to the standard for heart rate monitors, the implementation could also be used for other heart rate monitors which support the heart rate profile, for example the Polar H7 or Sigma R1.

7.4. Data Processing

To calculate the values for valence, arousal and mental effort, the physiological signals have to be processed in several ways. In the following the implemented processing steps are described. Besides calculating of these values, a configuration and baseline phase has been implemented, which will be described in detail in the following subsection.

7.4.1. Baseline Measurement

A baseline is measured before the model can be used. Baseline should be measured in a period of rest in a sitting position before first usage. The size of a measurement windows needs to be a power of 2, as the fast Fourier transformation is used for calculation. Measurement takes 5 minutes and 20 seconds, which corresponds to 5 measurement windows of 64 seconds and can be manually started.

A function is implemented, that allows the user to reset the baseline and measure it again. From time to time or when sensors have been removed this function should be used.

Baseline is needed for normalization of EDA and the result of power spectrum density. For PSD five windows are calculated and the lowest is stored for later normalization. EDA uses the minimum value in expressed in µS for normalization. In both cases, if there is a lower value in future measurement, the current baseline is replaced by this value.

7.4.2. EDA Analysis

EDA is received in µS. Tonic level of EDA is analyzed, based on the change in comparison to the baseline and the current maximum value. EDA is normalized with the simple formula:

$$SCL_{norm}(t) = \frac{SCL(t) - SCL_{min}}{SCL_{max} - SCL_{min}} * 100 \tag{7.1}$$

Maximum value is stored and replaced, if the current value is higher than the maximum value. The longer the measurement period is, the more accurate is the normalization. Within the first time of measurement it might not be reasonable to use the normalized data but the normalized change in percent between baseline and current value. Another option would be to let the user experience a situation of high arousal after baseline measurement to also get an estimation for the maximum value.

7.4.3. Power Spectrum Density

To calculate the power spectrum density of Heart Rate Variability (HRV), several step have to be done. At first, data needs to be preprocessed. In the next step, data is divided into several windows of 32 or 64 seconds (depending on configuration) length to apply a sliding window approach. Length of these windows has been chosen based on the further processing steps and on the fact that the reaction time should not be too long in mobile environments.

After the preprocessing, a Fast Fourier Transformation (FFT) is applied to the windows. FFT offers the advantage of a faster and more efficient processing of the data as the normal Fourier transformation. One requirement of the FFT is, that the length of the input has to be a power of 2. As there are no integrated functions for signal processing in Java a library [SW15] has been integrated for several signal processing functions, e.g. the FFT.

The output is the value in µS of the low frequency area between 0.06 Hz and 0.14 Hz. To determine the mental effort of a user, this value has to be normalized with the average of the baseline measurement. Result is a number between 0 and 1, where 0 means that mental effort is low and 1 stands for a high mental load.

$$mentalLoad(t) = 1 - \frac{PSD(t)}{PSD_{baseline}} \tag{7.2}$$

Calculation of PSD was evaluated by comparing the results of different data sets with other established programs, like Kubios HRV [TNL+14]. For this reason, a set of measurements were done with Polar RS800. Measurements were saved as text files which were used as an input for both programs. Results showed that the results of calculation were similar.

7.5. Integration of Context Information

Android offers access to many of the integrated sensors of a smartphone. In this work, context information based on information from the smartphone itself, concentrated on motion. The implementation of this aspect into the background service is further explained in the following subsection. Context information transmitted by the application itself is explained in the last subsection.

7.5.1. Motion

Different kinds of motion can be registered. On the one hand motion of the smartphone itself, on the other hand motion in the environment. The second one is the more interesting aspect for this work. Most smartphones with Android offer different ways to collect data about motion of an user. Gyroscope and accelerometer are hardware based sensors, which are integrated in modern smartphones. Gravity, linear acceleration and rotation vector sensor can exist as hardware or software solution in a smartphone.

Android offers an integrated API for a Step Sensor. Two different kinds are available, the step counter sensor and the step detector sensor. Step counter sensor has a latency of up to 10 seconds, step detector has only 2 seconds latency. The first one offers more accuracy. In this work, the more accurate step counter sensor is used, due to the fact that data is transmitted continuously and changes in adaptation might be fast enough.

Once initialized, the step sensor is triggered after each recognized step. The return value are the amount of steps taken since initialization. Step detector would only trigger an event and return a timestamp than the total amount of steps taken.

Both step sensors need a special hardware solution. Nexus 5 is one of the few devices at time of implementation, which offered this solution. Step sensor is available since API level 19 of Android and Android version 4.4.

7.5.2. Information from Applications

Besides information from the device itself, information can also be received by the application. Several information about performance or usage might be interesting for adaptation, based on the kind of application. This information is optional and not necessarily to be transmitted by the application.

The background service has the possibility to receive a value from the application once background service and application are connected. This value is a number between 0 and 1 and expresses how well or successful a user is doing. The range of values is split into five different zones, standing for ”very negative”, ”negative”, ”neutral”, ”positive” and ”very positive”. How the statistics of the application are mapped to these five states is decided by the application itself.

7.6. Integration of Fuzzy Logic

Different frameworks and libraries exist written in Java, offering fuzzy logic functionalities. In this work jFuzzyLogic [CAF13] has been used, an open source Java based library, offering a broad range of membership functions and defuzzifier. jFuzzyLogic is offered in two different versions, a full version and a smaller core version. For integration into the MUSE engine, the core version has been used as it offers a higher performance with limited resources and visualization is not needed. For definition of fuzzy rules, the standardized Fuzzy Control Language (FCL) is used.

The rules are defined in the Fuzzy Control Language in a separate file. This allows changes and adaptations of the rule set without knowledge of Java or looking into the source code of the engine. For realization of the fuzzy logic rules described in the chapter 6, separate files for the first fuzzy system, transforming physiological signals into affective and cognitive factors as well as a separate file for affective and cognitive state have been created.

Membership functions can easily be described (see 6). In a similar way fuzzy variables and defuzzification can be described. For more details see [CAF13].

```
FUZZIFY service
    TERM poor := (0, 1) (4, 0) ;
    TERM good := (1, 0) (4,1) (6,1) (9,0);
    TERM excellent := (6, 0) (9, 1);
END_FUZZIFY
```

Listing 7.1: Example for membership function in fuzzy control language

7.7. Connection to Background Service

Several functions and steps need to be implemented in the application to receive the results from the background service. After receiving this information, the application decides which values are used and which adaptation takes place under which circumstances. In the following the most important steps for integration into an application are introduced.

7.7.1. Binding the Service

When the background service is running, applications can connect easily to the service. The application needs an Activity that handles connection and received values for adaptation. Multiple applications can bind to the service at the same time.

7.7.2. Managing Service Lifecycle

To handle and manage the service lifecycle, the service connection used in bind service needs to be implemented. The service connection initializes the connection, as soon as the service is bound and requests data from the service. Service connection handles also the case, if service gets disconnected. The following shows an example for the Service connection:

```
private final ServiceConnection mConnection =
    new ServiceConnection() {
    @Override
    public void onServiceConnected(ComponentName componentName,
    IBinder service) {
        mService = new Messenger(service);
        mBound = true;
        try {
            Message msg = Message.obtain(null, 1);
            msg.replyTo = mMessenger;
            mService.send(msg);
        } catch (RemoteException e) {}
        Message msg = Message.obtain(null, MSG_GET_HRV);
        msg.replyTo = mMessenger;
        try {
            mService.send(msg);
        } catch (RemoteException e) {
        e.printStackTrace();
    }
    }
    @Override
    public void onServiceDisconnected(ComponentName
    componentName){}
};
```

Listing 7.2: Example for service connection

7.7.3. Implementing Incoming Handler

Background service needs an instance of IncomingHanlder to handle the received messages. Several messages are defined for the single values. These, and their possible values are listed later in this subsection. The following code shows the Incoming Handler:

7.7.4. Additional configuration

For security reasons, Intent Filter and service need to be added in the AndroidManifest.xml of the application. The following shows the lines added to one of the example applications:

```
<intent-filter>
    <action android:name="android.intent.action.MAIN" />
    <category android:name="android.intent.category.LAUNCHER" />
</intent-filter>
```

Listing 7.3: Configuration of Intent Filter for connecting applications

Without these lines added, Android blocks the connection between application and service.

7.7.5. Transmitted Values

The adaptation itself is done by the application to allow a broad range of different kinds of applications. Based on the given values, the application has to make a decision if and which elements get adapted. Based on the kind of application, some values might be ignored. Transmitted values and the possible values are listed in the following table:

Value	Possible Values
alarmed	high, medium, low, very low
excited	high, medium, low
happy,content	high, medium, low, very low
relaxed	high, medium, low, very low
tired	high, medium, low, very low
bored	high, medium, low, very low
sad	high, medium, low, very low
frustrated	high, medium, low, very low
cognitive state	high, medium, low, very low
movement	high, medium, low
cognitive state	very high, high, medium, low

Table 7.1.: Transmitted values of the engine

The values are all defined in the model and concept, introduced in chapter 6. The first eight values are from the affective state. Values for cognitive state are listed at the bottom.

Besides these values, the raw data is also transmitted. This makes sense for applications, where physiological data is used for direct adaptation and control. For example to control speed of gameplay based on heart rate.

7.8. Configuration and Start of Service

The background service offers a simple user interface for configuration and start of the service. In this user interface, it can be chosen, which sensors should be used. The interface (shown in figure 7.3) is very simple designed and offers a quick view of the current measured values. The corresponding affective or cognitive State is not visualized at the moment.

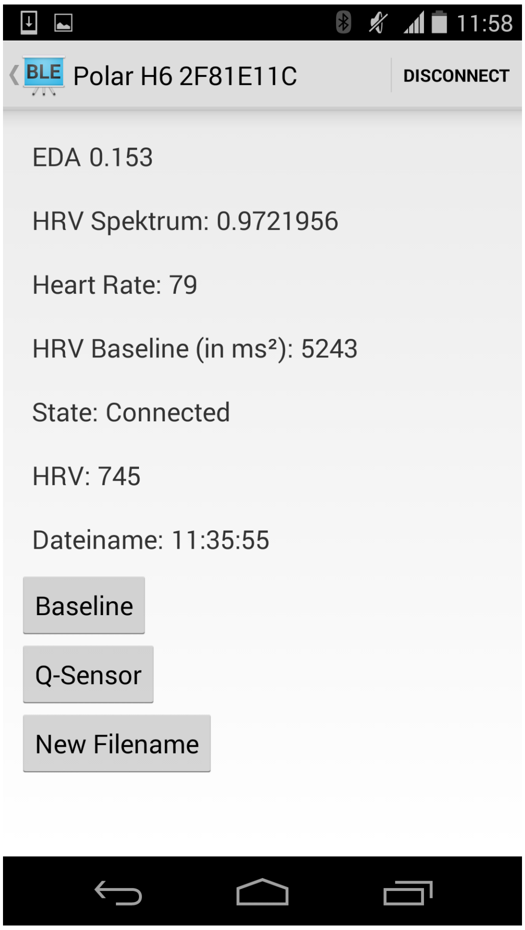


Figure 7.3.: User Interface of the background service, overview of current measures

The overview allows to control if the signals are connected and that the values are in a reasonable range. EDA, HRV spectrum, HR, HRV Baseline and HRV are shown, together with buttons allowing to restart baseline measurement or to connect/disconnect sensors.

After start the Bluetooth connection is checked and a list of available sensors is shown. The user can choose between these sensors. If Bluetooth is not activated, a dialogue opens which asks to activate Bluetooth. Once sensors have been chosen, the overview screen is shown. After starting the service, it is running in the background until terminated completely by the user.

8. Applications and Adaptations

As described in chapter 5, different application areas and scenarios might take advantage of the current user state. In the scope of this work, several applications have been developed as a proof of concept in the area of rendering, games, eLearning and adaptation manager, addressing different concepts for use of MUSE engine. The applications, the integration of the engine and the results of first studies are explained in this chapter.

8.1. Zone of Impulse

"Zone of Impulse" is a multiplayer space-shooter for smartphones, that was developed at Goethe University Frankfurt [RSK12] [Rei11]. The goal was to develop a game, that allows two players with different skill levels to play together. If one of the players is overstrained, gameplay adapts to make it easier for this player. On the other side, gameplay gets harder for a player if the player is not challenged enough. The biofeedback version adapts game difficulty in a way that allows both players to enjoy gameplay, independent of their skills.

8.1.1. Application

The goal of the game is simple: two players compete against each other and have to try to shoot the opponent as often as possible within a limited time frame. The game is settled in a space environment. The interface is shown in figure 8.1. Each player controls a spaceship by using the tilt sensor in the smartphone, shooting is done by tapping on the screen. Further controls are shown in figure 8.2. The score is shown next to the health bar in the upper left and upper right corner of the screen. In the middle of the upper screen the remaining time is shown.

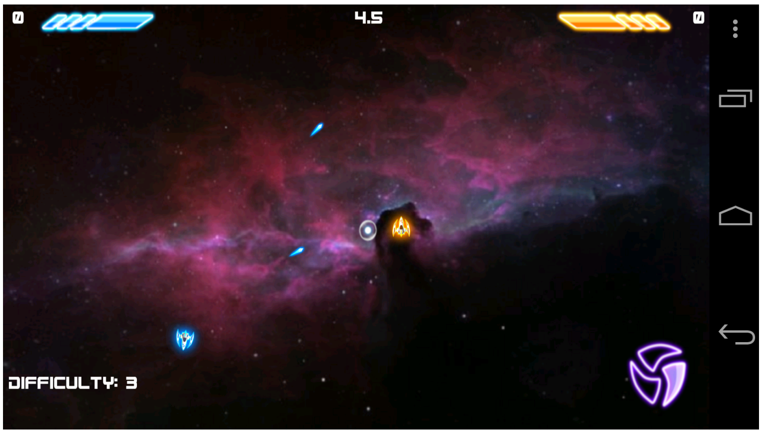


Figure 8.1.: Screenshot of the extended version of Zone of Impulse

In the lower right an icon visualizes the progress of the special ability slow-mo. The slow-mo ability can be activated, when the icon is fully charged. After activating, the spaceship of the opponent is slowed down for a few seconds. Another special ability is the shield which can spawn in the center of the screen. This ability protects the player for several seconds from getting shot by the enemy.

Originally, electrodermal activity (EDA) and heart rate (HR) were measured to adapt several game elements. Electrodermal activity and heart rate were collected by a stationary device in the original version. Gameplay is adapted based on the score and the interpretation of the physiological data. Game elements that can be adapted are speed of the own ship, recharging time of the slow-mo ability and shield spawn time.

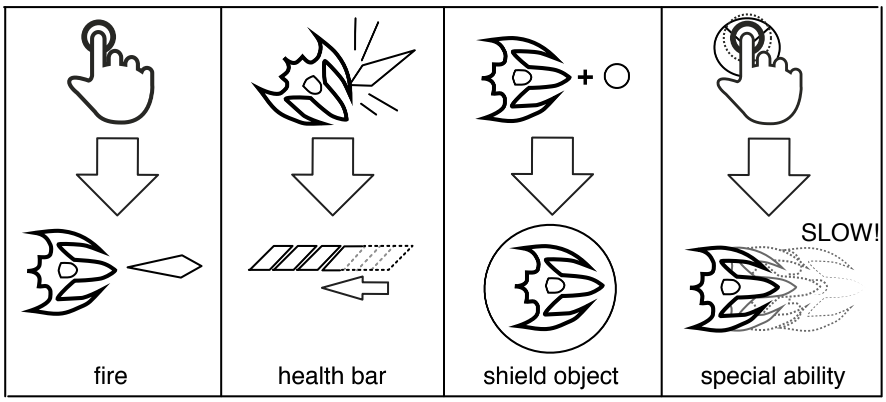


Figure 8.2.: Control of gameplay [Rei11]

The original version developed in the master thesis [Rei11] was extended by several items and instead of EDA and heart rate, the MUSE engine was integrated. In the extended version a visualization of the difficulty level is integrated in the interface, shown in the lower left of the screen. One big aspect that changed were the sensors for measuring EDA and heart rate, which are now smaller and more suitable for mobile scenarios as the previous sensors. The changes of the adaptation rules by integrating MUSE engine will be described in the following subsections.

8.1.2. Model

For adaptation of gameplay, only affective state has been used combined with the current score of the player. The eight different affective states of the MUSE engine have been divided into desired, undesired and transition states, like shown in figure 8.3. As it is the aim of the game that the user has a good time, the desired states (marked in green) are characterized by positive valence and a certain level of arousal. The two neighbored states are transition states (marked in yellow), where slight adaptations are done to get the user back in one of the green states. Undesired states are marked red and are characterized by a mainly negative valence. Adaptations in gameplay are done to get the user back in one of the desired states.

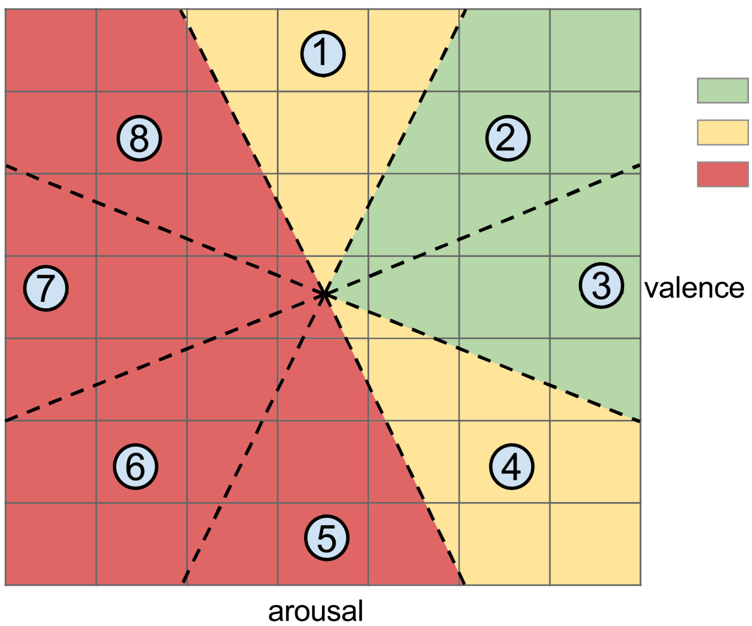


Figure 8.3.: Desired and undesired states during gameplay of Zone of Impulse

8.1.3. Rule Set

After integration of the MUSE engine, these rules were modified like shown in table 8.1. When the user is in a desired state, no adaptation is done. Adaptations are only done, if the user is in a transition or undesired state. In an undesired state, difficulty is decreased or increased more than in a transition state to get the user back into one of the green states faster. Within the game, user state is checked every 30 seconds to decide on possible adaptations. Initial set of adaptations and rules was defined based on previous studies.

Affective State	Score	->	Ship Speed	Shield	Slow-Mo
undesired state	high		decrease	2x decrease	2x decrease
undesired state	low		increase	2x increase	2x increase
transition state	high		no change	decrease	decrease
transition state	low		no change	increase	increase

Table 8.1.: Adaptation rules for Zone of Impulse

A high score means, that the player is currently leading whereas a low score means, that the player is currently loosing. A high score leads to an increase in game difficulty, where as a low score leads to an decrease of difficulty. Depending on the affective state, adaptation is done stronger (undesired state) or only slightly (transition state). Every adaptation aspect had five possible levels. Shield spawn time and recharge time of slow-mo ability are adapted in undesired states. If maximum level was reached, no further adaptation is done.

8.2. Beats Down

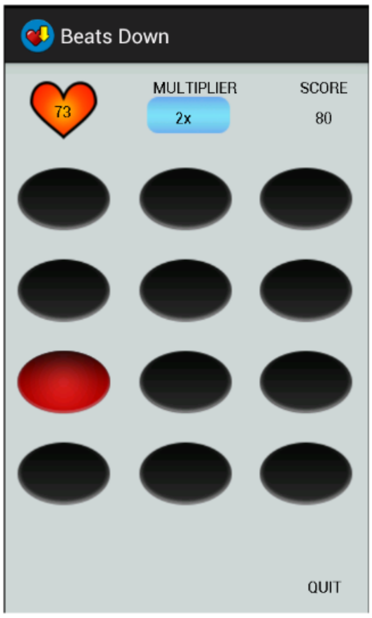
The game "Zone of Impulse" uses calculated values for affective and cognitive state for adaptation of gameplay. Adaptations were passively, instead of controlling gameplay actively. A further developed game "Beats Down" uses the heart rate for direct interaction with the game and addresses other aspects as "Zone of Impulse". Heart rate can be influenced directly by the user and can be increased or decreased. This aspect has been used, to implement two different game modes.

8.2.1. Application

The game was developed in a Bachelor thesis at Goethe University Frankfurt [Smy12]. The game is based on a simple whack-a-mole principle. The game field consists of 12 different fields (see figure 8.4(a)). If one of the fields is flashing up red, the user has to hit this field as fast as possible to collect points. Two different gaming modes were implemented: a challenge and a relax mode. In both scenarios, heart rate was used to influence gameplay directly.

In the challenge mode, users had to raise their heart rate to raise the frequency of blinking fields. Due to a higher frequency, more points can be achieved as there are more opportunities to hit the field. Raising of heart rate can for example be achieved by movement or sports. During a first study users did activities like climbing stairs and running around to raise the heart rate. In the relax mode, users had to calm down and lower the heart rate to get a bonus multiplier. Decreasing heart rate can be achieved by sitting down, relaxing or controlled breathing. Based on how much the heart rate decreased, multiplier was 2x, 3x or 4x.

The figures show the relax mode (figure 8.4 (a)), as well as a person playing the game during challenge mode (figure 8.4 (b)). The interface of challenge mode is similar to the relax mode.



(a) Relax mode

(b) Participant during gameplay

Figure 8.4.: Screenshots of Beats Down and photos of gameplay

In a first user study 13 users participated. All participants had to play each game mode in randomized order. After each gaming session, participants had to rate the game modes regarding different aspect, for example enjoyment during gameplay. The results showed that participants rated enjoyment in relax and challenge mode higher than the normal mode, which did not offer any physiological interaction.

8.2.2. Model

In this application, only raw values of heart rate have been used instead of affective and cognitive state. A 60 second long baseline has been collected right before the start of the game instead of the baseline collected by the background service. This has been done to ensure that the game works with an actual value instead of a baseline, that might have been taken in a sitting position and user is for example now standing.

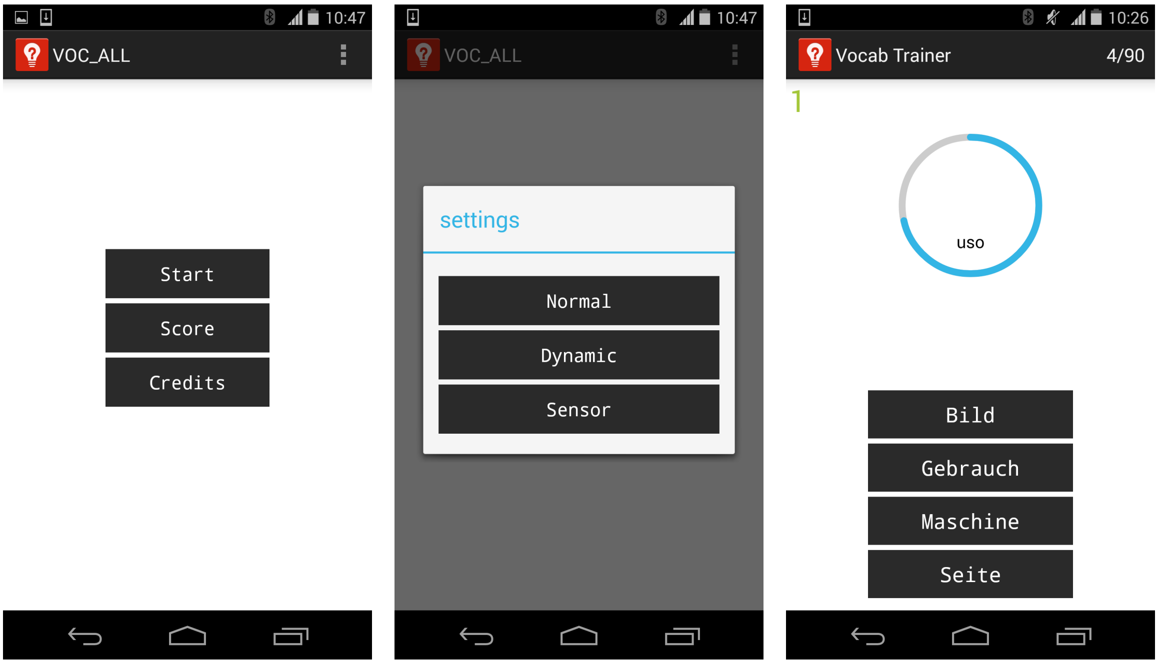
Game elements were adapted based on change in heart rate regarding the measured baseline. If current heart rate was 10 or 20 percent higher or lower than the previous heart rate, the different elements got adapted. For challenge mode, game speed is adapted by increasing the frequency of flashing lights by the percentage of heart rate raise. Relax mode on the other hand adapted the multiplier based on the percentage heart rate was decreased. The multiplier for doubling points (2x) was activated after lowering the heart rate 10%, 4x multiplier was activated for a decrease of 20%.

8.3. Affective Vocable Trainer

The vocable trainer was developed in a bachelor thesis [Kle13]. The goal was to develop a vocable trainer for mobile situations, where difficulty of vocables is adapted by the state of the user to support the user in the learning flow.

8.3.1. Application

The application is kept simple and offers 3 different learning modes (figure 8.5 (b)): normal mode without biofeedback, a dynamic mode where adaptation is based on performance only and a biofeedback mode. For adaptation based on physiological data, mental load and arousal have been chosen in the original version of the trainer. In the original application, baseline of Heart Rate Variability (HRV) and EDA was measured within a 5 minutes time frame before start of the game. After integration of the MUSE engine, rules were adapted and will be presented in subchapter 8.3.3.



(a) Main menu

(b) Setting window, showing three different modes

(c) Training of vocabulary

Figure 8.5.: Screenshots of the vocabulary trainer application

The different modes are kept simple. A word is presented together with different possible answers. Time is limited and visualized by a circle filling up (figure 8.5(c)). Some gamification elements like a score and bonus points when a certain amount of vocables in a row is correct are integrated. Currently only german-spanish is available as learnable language.

In a first evaluation during the bachelor thesis the three different modes (dynamic, sensor and normal) were evaluated regarding learning performance. In a study, 12 participants had to test each mode in randomized order. In the sensor mode, participants had more vocabulary correct and a higher score (72 of 90) than in the normal (60 of 90) and dynamic mode (67 of 90).

8.3.2. Model

Originally a simple rule based system was integrated for adaptation of difficulty level. With integration of MUSE engine, arousal and mental load were mapped to affective and cognitive state. By integrating the engine, baseline values are submitted directly. Baseline measurement can be removed from the application, allowing a direct start of the game.

Thinking in regards of the affective state, the green marked states in figure 8.6 have been chosen as positive for learning. To achieve these states or to keep an user in one of these states is the goal. Desired states are characterized by a medium cognitive state and affective states with a certain level of arousal and positive valence. The red marked states are critical and it is not desired that the user is in one of these states. The undesired states can be identified by a low arousal and negative valence. This also takes into account the Yerkes-Dodson curve. In these states, users are not activated and tend to have a negative valence. To reach a certain performance during the learning process users need to be activated.

Yellow states are states, which are between desired and undesired states. In this case, it is not critical to get the user immediately into a green state but first steps are taken to get the user back in a green state.

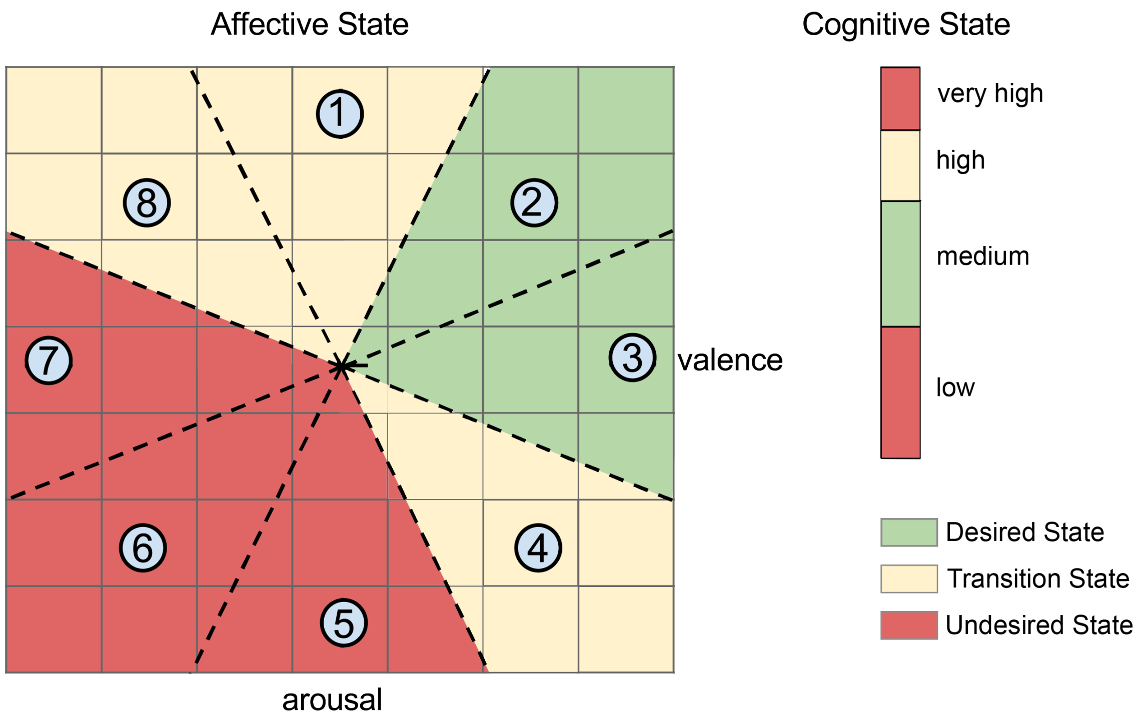


Figure 8.6.: Desired and undesired cognitive and affective states during usage of the vocabulary trainer

Regarding cognitive state very high and low values are undesired states. If mental load is low, difficulty of vocab trainer is raised. On the other hand, a very high mental load leads to a decrease in difficulty. In total the application supported three difficulty levels.

8.3.3. Rule Set

Additional context information besides performance is integrated. In this scenario, steps of an user are used to identify movement. To adapt difficulty, timer and difficulty of the vocabulary can be changed. Each aspect has three different levels. The rule set was extended from 4 to 15 different rules. Table 8.2 shows the rules, explaining which states lead to an increase or decrease in timer speed or difficulty. If the user is already in a desired state, no further adaptation is done. The rules handle only the situations in which an adaptation should be done.

Affective State	Cognitive State	Steps	->	Timer	Difficulty
undesired state	very high	low OR high		decrease	decrease
undesired state	high	low		decrease	no change
undesired state	low	low OR high		increase	increase
undesired state	medium	low		increase	no change
undesired state	high	high		decrease	decrease
undesired state	medium	high		decrease	decrease
transition state	low	low		increase	increase
transition state	medium	low		increase	no change
transition state	high	low		decrease	no change
transition state	very high	low		no change	decrease
transition state	low	high		no change	increase
transition state	medium	high		increase	no change
transition state	high	high		decrease	no change
transition state	very high	high		decrease	decrease

Table 8.2.: Adaptation rules for vocabulary trainer

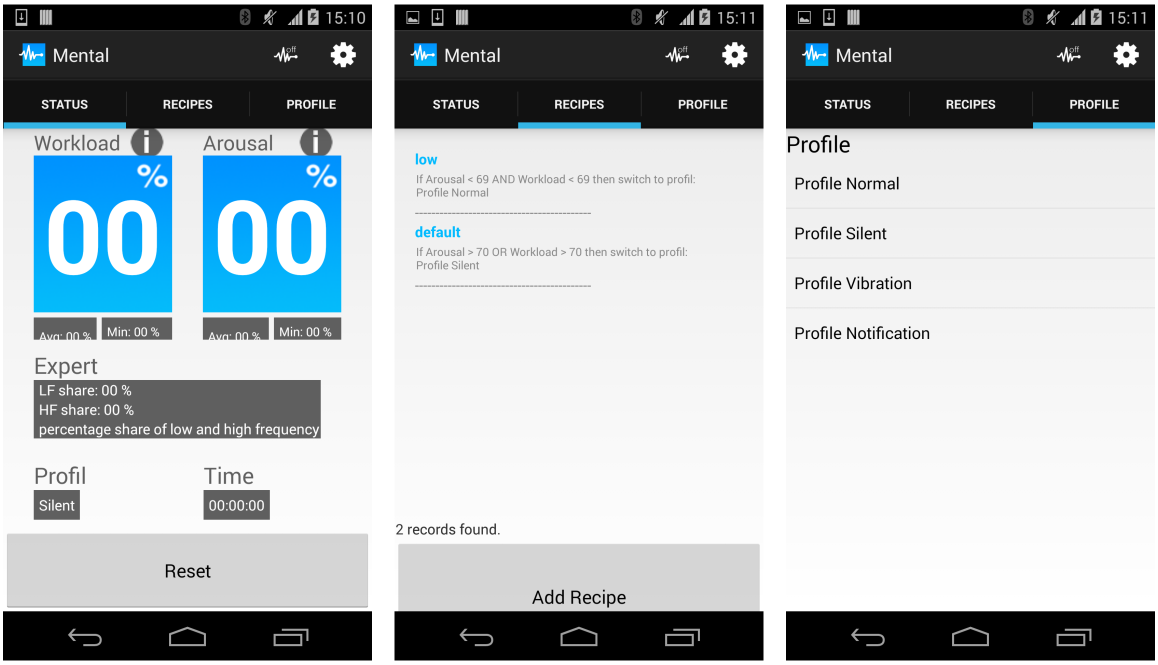
The application checks every 60 seconds for adaptation to prevent a too fast adaptation which might lead to a negative state.

8.4. Mental Adaptation Manager

The application Mental was originally developed in a bachelor thesis [Reh13]. Mental offers an overview of physiological measures of the user, which can be used to configure adaptations. In the adaptation manager, users can configure several so called recipes, which allow to adapt different basic functions of the smartphone. For example, if the user is concentrated (expressed by a high mental load) and should not be disrupted, the ringing tone can be set to silent.

8.4.1. Application

In the main window of the application, all physiological signals that are connected to the application are visualized (figure 8.7 (a)). Workload and Arousal are normalized based on the baseline value and the highest value measured in the measuring window and are visualized as a value between 0 and 100. Some other aspects are also shown, like details of the workload measurement and the actual EDA value. Figure 8.7 (b) shows the tab in which the recipes can be configured. In a simple "IF... THEN..." style rules can be configured in the application to switch to different profiles, e.g. work or home. The profiles themselves can be configured in the third tab. Currently functions like volume of the ringing tone, rejection of calls and notification about new messages can be configured.



(a) Overview of workload and arousal in the status tab

(b) Overview of all configured recipes

(c) Configuration of different profiles for adaptation

Figure 8.7.: Screenshots of the application Mental

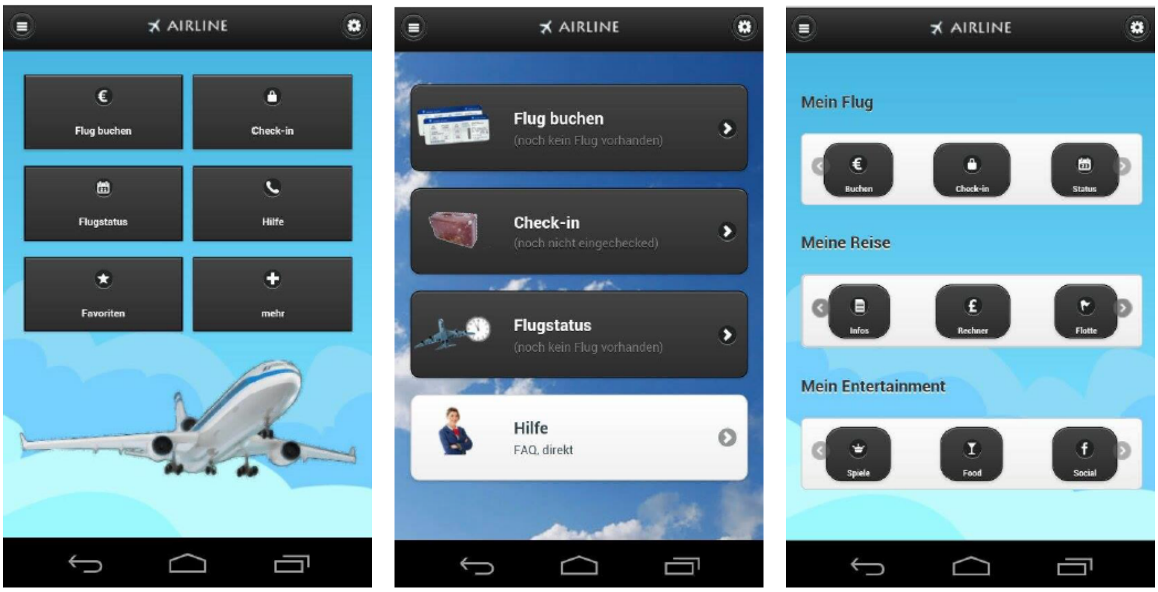
When giving the configuration of adaptation rules into the hands of the user, the design of the application has to make sure, that the values are understandable for the user. A first study solely concentrated on the usability and understandability of the interface itself, instead of testing the whole application. For user with a deeper knowledge of physiological signals, a window with additional values is integrated.

8.4.2. Model

The model is not applied itself as the rules are made by the users themselves. The model is only used to correct and control physiological data. The application is more a way of configuring the rules for several adaptations of basic functions. At the moment the application is working with the normalized values, but an extension for an integration of a visualization of the model is planned.

8.5. Airline Application

The airline application was developed in a master thesis [Kol14]. The goal was do develop different interfaces for different situations to examine if adaptation of user interface elements might be useful depending on the situation. In the application, three different interfaces are offered based on the state of the user. At an airport, users can be under stress because of catching a flight in very short time. On the other hand, users at an airport can be bored because of long waiting periods for the next or a delayed flight. The different states covered by the interfaces are boredom, neutral and stressed.



(a) Mixed interface (b) Reduced interface (c) Extended interface

Figure 8.8.: Screenshots showing the main menu in the three different modes of the airline application

8.5.1. Application

To address the three different states, three interfaces were developed: reduced, mixed and an extended interface. Depending on which user interface is shown, different functionalities are offered. The reduced interface, which is offered to user in very stressful situations, offers a reduced interface (shown in figure 8.8 (b)) with only the most important information like check-in, buying tickets and flight schedule. In neutral situations, where the user is whether stressed nor bored, a mixed interface is used. The mixed interface, shown in figure 8.8 (a), offers more functionality than the reduced interface, like an option to mark favorites and more functionalities hidden behind the "more" button. The extended interface is designed for situations where users are bored and offers plenty of functionalities. The interface is shown in figure 8.8 (c). Besides the function that are offered by the mixed interface, the extended interface additionally offers tools like a currency converter or entertainment like games.

Besides the main menu and the interface, some functionalities also differ in the different versions. One example is the booking of a flight. In the reduced mode, passenger information is already filled out in the form and no additional information about the destination is shown. In the extended interface, passenger information is not automatically filled out, but additional information about flight and destination are shown.

In the original version of the airport application, the MUSE engine was not integrated directly but used for testing. The user interfaces were only used for testing and evaluating them under different stressful situations. The integration of the MUSE engine allows live adaptation of the interface at start of the application. Adaptation did only take place at the start of the application and not during usage to minimize irritation of the user.

8.5.2. Model

Depending on affective state and cognitive state, an interface is chosen. As shown in figure 8.9, the dark blue marked states are states of high stress or frustration, where the reduced interface should be used. The medium blue marked states, state 2, 3 and 4 are states with a positive valence and a balanced arousal level, which would be optimal for the mixed user interface. States 5 and 6, marked in light blue, would be the states where the extended interface should be preferred as the user is in an inactive state.

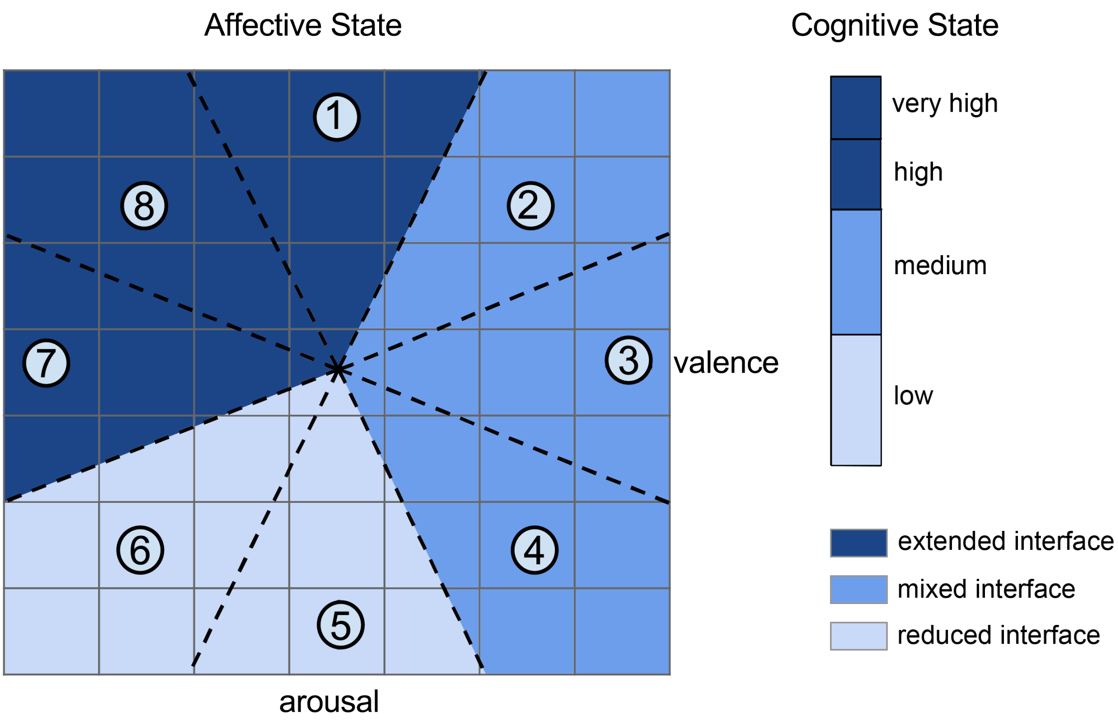


Figure 8.9.: Assignment of the three interfaces to affective and cognitive states

When the user is in one of the inactive states and/or cognitive state is low, the extended interface is offered that gives more information and functionalities like games to activate and involve the user. In stressful states and/or if cognitive state is very high or high, the reduced interface is offered to not stress the user any further than necessary and to support the user. Medium cognitive load or balanced affective states lead to the mixed interface.

8.6. Adaptation of Level of Detail

Besides the examples for mobile applications, the MUSE engine was also used for testing of level of detail (LOD) algorithms aiming for a LOD adaptation based on physiological signals. The level of detail influences the quality of a scene. If a user is less attentive, lower quality in the rendering of images may be sufficient. This may especially lead to a better performance on mobile devices by applications using this adaptation mechanism. In the past, eye tracking has been used in different research projects [CCW03] to adapt the (LOD) of a rendered scene to speed up rendering. Based on the arousal of a user, conclusions regarding attention can be drawn in a more general way, than in case of e.g. eye-tracking.

8.6.1. Concept

The goal is, to select a LOD algorithm based on the EDA of a user. The workflow of the concept is shown in figure 8.10). Based on EDA, a LOD algorithm is selected. In our first evaluation, two different kinds of LOD algorithms have been tested. In the next step, rendering of the scene is adapted and presented to the user.

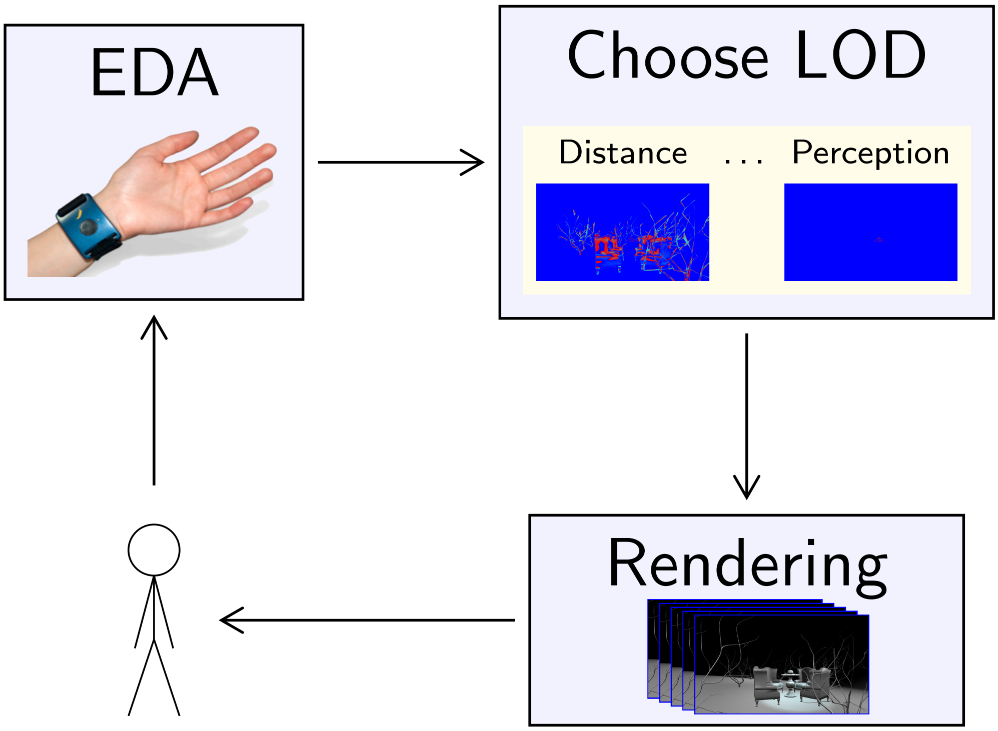


Figure 8.10.: Workflow of LOD adaptation based on electrodermal activity

The first study aimed for a comparison of different LOD strategies and detail levels in pre-rendered videos. Figure 8.11 shows an example of two different LOD versions of a scene. The images with the blue background show the difference of the current rendered image in comparison to the original image. If EDA is high (left part of the figure), a higher LOD is assumed to be more suitable. In the other case, if EDA is low, a lower LOD is sufficient.

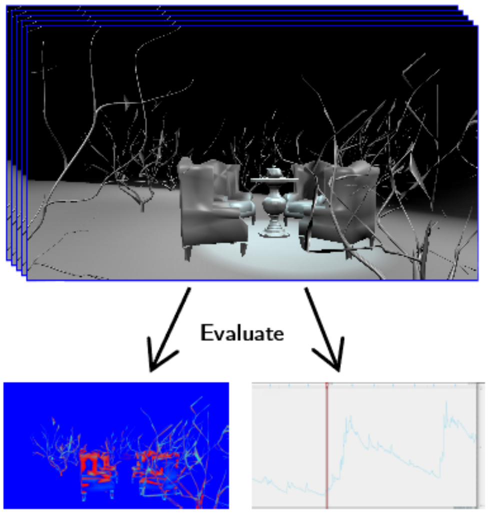


Figure 8.11.: Example scenes of the test videos with different LOD and EDA curves

8.7. Recommendations for Integration of MUSE

Different types of applications have varying requirements in details of the user state. In this chapter, different examples and areas for applications were presented. As an example, learning applications might profit more from cognitive state where on the other hand gaming applications might profit more from affective state. Therefore it is crucial to analyze the requirements of applications before designing adaptations or using it for other purposes. In the following, the presented applications are summarized and recommendations given.

8.7.1. Gaming & Entertainment

Two different games have been used, to test the integration of the MUSE engine. The first example, Zone of Impulse (see section 8.1) is a game, which aims to keep the user in a flow state (2.7.5) by using the user state indirectly. Preferred states were reduced to two of the eight states of the model (state 2 and 3), whereas state 1 and 4 were states, where the game reacted slightly to get the user back into one of the desired states. The remaining states are not desired and lead to a stronger reaction of the game, to get the user in one of the desired states.

To test direct control of gameplay by physiological data, the game Beats Down, presented in chapter 8.2 has been developed. First tests showed, that usage of affective or cognitive state are not suitable for direct control, as participants were not sure about how to control them. Therefore, physiological signals were integrated without the interpretation of affective or cognitive state. Based on the test results, we recommend to select and use physiological signals for direct control carefully. In related work of Nacke et al. [NKLM11], results were similar. It should be clearly understandable for users, how the signals can be influenced. For most users it might be clear, how to influence heart rate but for example not how to influence electrodermal activity. A visualization of the physiological values in the interface of the game supported the user in the task.

8.7.2. Other Applications

Besides gaming and entertainment, three other mobile applications have been developed using cognitive and affective state. The first example was a vocable trainer application presented in 8.3, adapting different parameters like difficulty of vocabulary or speed. As the application is focused on learning, cognitive state was mainly used for adaptation. Besides cognitive state, affective state was also used to ensure that the user is not getting into a demotivated or frustrated state. First results showed, that participants were more effectively learning in a version with adaptation.

In the second example, an airport application (see section 8.5), the interface was adapted. Depending on the affective and cognitive state, one of three different user interfaces is displayed. User interfaces differed in complexity of the shown options and in visual aspects and were optimized for stressful, boring and neutral situations.

The third application, the application Mental (section 8.4) offers to define adaptation rules for different functionalities of mobile devices, like e.g. muting the ring phone.

8.7.3. Other Examples

Besides mobile applications, the MUSE engine has also been used in testing of other applications. One example, which was presented in this chapter, is the concept for adaptation of LOD 8.6. In a first study, different videos were presented to measure the impact of a searching task regarding level of detail to physiological signals like electrodermal activity. Results showed, that it might be possible to adapt level of detail based on the electrodermal activity of an user. This offers a broad spectrum of possible applications, especially adapted rendering on mobile devices to save resources.

9. Evaluation

After detailed description of the concepts and its applications in the previous chapters, the different theses from chapter 4 are further investigated with help of several studies. Before the evaluation of the developed model and theses, the signal processing for determination of congitive and affective state are investigated. At first, two studies for evaluation of cognitive state estimation will be presented, followed by a first study for the evaluation of affective state estimation.

Finally, a study is presented addressing the different theses. With regards of the described theses, different aspects of the model are evaluated and presented in detail, like usage of the model in outdoor scenarios or loss of input channel.

9.1. Prestudy - Evaluation of Cognitive State

To evaluate the estimation of cognitive state its main component, the determination of mental load based on heart rate variability has been evaluated in two different studies. Both studies and its result will be described briefly in the following.

9.1.1. Study I - Measuring Mental Load with a Polar Heart Rate Monitor

In a first study presented in [SK11], the results of spectral analysis were compared to subjective ratings of NASA Task Load Index (NASA-TLX) rating, to determine if the perceived subjective workload correlates to the interpretation.

Four different tasks had to be solved by the participants, which were expected to trigger different levels of mental load. Task 1 was a simple reading task, where participants had to read a short simple text within 4 minutes. In task 2, participants had to solve verbal arithmetical tasks while sitting. Task 3 consisted of a Stroop Color Test [Str35], whereas task 4 was similar to task 2 but involved walking. In total 6 male participated with a mean age of 28 years.

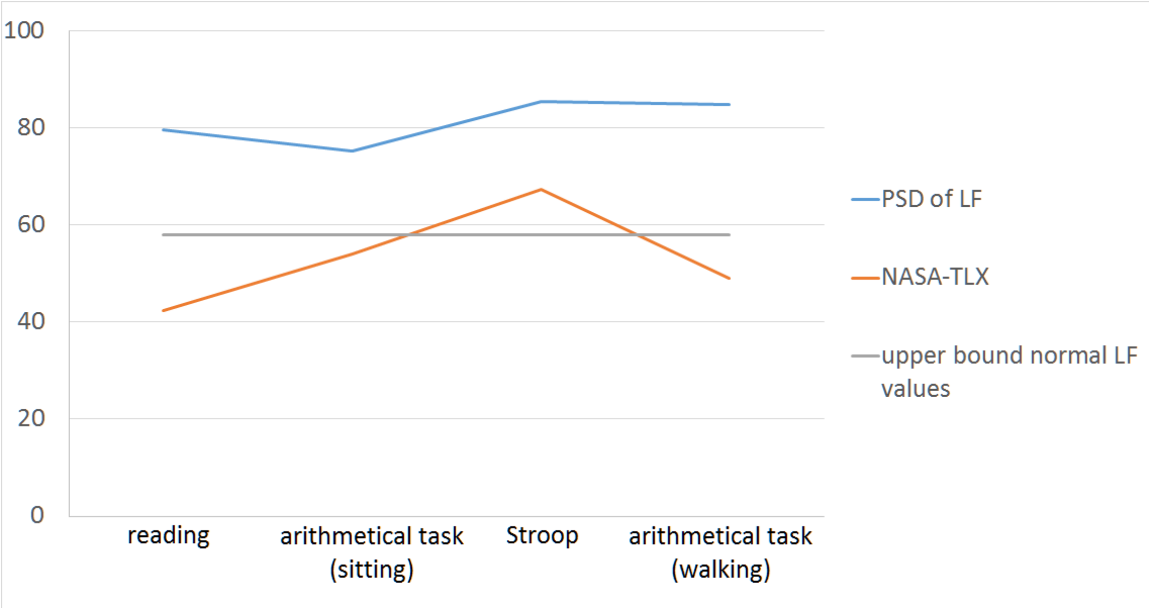


Figure 9.1.: Normalized results of the study [SK11]

The results showed, that mental load values for the tasks differed. Figure 9.1 shows the results of NASA-TLX rating, result of spectral analysis as well as the upper bound for normal low frequency values of spectral analysis as found by Malik et al. [MBC+96]. Analysing the correlation between PSD and NASA-TLX results of the tasks, a correlation coefficient of 0.34 is returned. For 6 participants this value does not indicate a significant correlation. Task 1 had the lowest results in the subjective rating and also low results in the results of spectral analysis. Task 2 had even lower results regarding PSD analysis, even if it was rated as more stressful than task 1 in the NASA-TLX rating. This could be due to the design of task 1. During task 1, participants were under a high amount of stress at the beginning, but then recognized that they had more then enough time at the end of the task. Task 3 and 4 had nearly the same results in the spectral analysis, but task 4 was rated lower in the NASA-TLX. This might be influenced by the walking during task 4.

To summarize the results, the most stressful task according to PSD analysis, was also the most stressful task according to subjective rating in the NASA-TLX. The results gave a first hint, that it could be possible to distinguish the mental load of different tasks by heart rate variability analysis. Furthermore, the results also showed, that control of movement might be crucial for correct interpretation of mental load, when looking at the results of task 4.

9.1.2. Study II - Visual Recognition Task

In a study, presented in [SSK11], the implementation of mental workload analysis was further investigated. Heart rate variability was measured for determination of mental workload, as well as NASA-TLX for a subjective rating of workload. The aim of the study was to find out, when a user recognizes different variations of an email icon. Variations of the email notification icon ranged from a small envelope in the task bar, to a more salient bigger blinking icon.

In total 10 participants took part, aged between 23 and 29. Participants had to solve three different tasks. Task 1 was a simple task of comparing the quality of images, task 2 was resting in front of the computer and task 3 solving Sudokus. The normalized results for all three tasks are shown in figure 9.2. Details of the study can be found [SSK11].

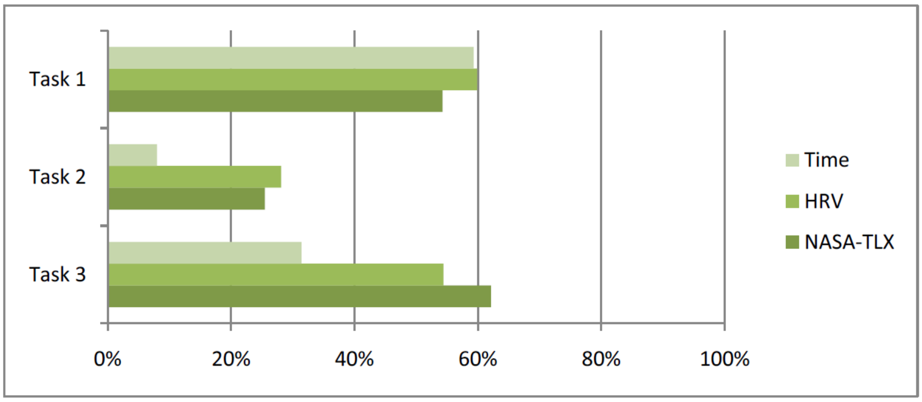


Figure 9.2.: Normalized results of the study

Time needed for first recognition of the icon as well as normalized (Heart Rate Variability (HRV)) and NASA-TLX results were compared. The results showed a correlation of estimated mental workload and subjective rating. Running a linear regression resolves in $r^2 = 0.88$. Further details of the study can be found in [SSK11].

9.1.3. Conclusion

The results were gained with tasks and activities addressing different level and aspects of mental workload. Both studies showed, that results of spectral analysis were comparable to the results of subjective ratings, which leads to the assumption that cognitive state results in the MUSE engine are reliable.

9.2. Prestudy - Evaluation of Affective State

In contrast to cognitive state, affective state is challenging to evaluate. As in some examples in the state-of-the-art section mentioned, it is difficult to measure the whole spectrum of affective states, especially the states with negative valence, e.g. sadness.

9.2.1. Evaluation of Arousal

For the example of combining measurement of arousal with adaptation of level of detail (LOD) presented in chapter 8.6, a first study was conducted. Electrodermal activity (EDA) was measured as well as participants had to fill out a SAM questionnaire (see chapter 2.5.2 and 2.6.2) for a subjective rating of their arousal and valence on a nine point scale.

In a first study, 12 videos with a length of 90 seconds were produced with two different scenes and two different LOD algorithms (distance and projection-size). For a better comparison of the results, a variant of the video without any applied LOD algorithm was produced. The aim of the study, was to examine if there is a correlation between the activation level of an user measured in EDA and the noticed differences by the user in the rendered scenes.

The study was divided into two sequences. In the first sequence, each participant was asked to watch six videos and rate the quality afterwards. In the second sequence, participants were also asked to especially look at the quality of single objects in the scene to get them more activated followed by the quality rating. The quality rating was divided into 10 items, where 1=very bad and 10= very good. The videos in each sequence were presented in randomized order. The results lead to the assumption, that it is possible to adequately adapt the LOD based on the current arousal level derived of the EDA.

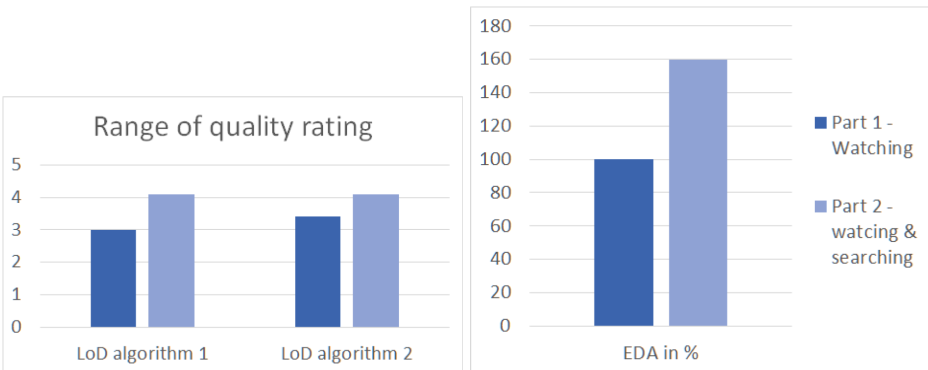


Figure 9.3.: Results for quality rating and EDA in the first study, comparing the first and second sequence

In total 24 participants, aged between 21 and 35, 5 female and 19 male, took part in the study. Figure 9.3, the results of the quality rating an EDA measurement are shown. The results of the quality rating (shown in the left of the figure), show that quality rating of the videos was broader in the second part. This leads to the observation, that participants rated the visual quality of a scene in more detail when they were more actively involved.

Looking at EDA results, shown in the right part of the results, 83.3 % of the participants showed a significantly higher skin conductance level in the second part of the test, when actively searching for differences.

Comparing the results of arousal rating in the self-assessment questionnaire manikin (SAM) scale with the results of EDA, no correlation could be found. This may be explained by that the subjective experienced arousal was not big enough for a difference in the ratings, but could be detected with EDA.

9.2.2. Conclusion

The results for evaluating the correctness of affective state are mixed. As stated, the evaluation of affective states or emotions is difficult, as in most studies the whole range of states cannot be covered. The results of evaluating arousal in 9.2.1 were not satisfying, which might be due to a failure in study design or lack of subjective rating scale.

9.3. Evaluation of Theses

A concept and model for the interpretation of physiological signals and context information to a user state has been developed in chapter 5 and 6. Applications of these models have been presented in chapter 8. To prove the three theses for this model introduced in chapter 4, different aspects need to be evaluated.

It is difficult, to compare the developed model with other models as introduced in related work (see chapter 3). The models differ in input channels and several aspects, that might lead to not comparable situations. Therefore it was decided instead of comparing to existing models, to compare the model of the MUSE engine under different situations. In the following, the concept for proving the three theses will be presented.

9.3.1. Combination of Physiological Signals and Context Information

Thesis I (see 4.2) investigates the combination of context information and physiological signals for improvement of user state interpretation. The implemented model offers two different ways of context information integration: for control and direct improvement of user state interpretation.

Context information for control of physiological signals is only important, if the user is outside of a controlled environment, where signals can be influenced by aspects of the environment. The second step, using context information for direct improvement of interpretation can also influence the interpretation in controlled environments.

To prove the thesis, the test needs to compare the full model with a version of the model, which only has physiological signals as input and a control group which has no integration of the model at all in the application. The test has to be taken in controlled environments as well as in outdoor environments, to evaluate the usefulness of context information. The results will be described in section 9.9.

9.3.2. Model for Mobile Scenarios

The model was developed for mobile scenarios, offering different improving and controlling aspects for mobility. To test thesis II (see 4.2), that a model for the integration of physiological signals and context information in mobile scenarios can be defined, tests need to be done outside the lab under realistic conditions of everyday life.

A test has been developed, involving different scenarios: test in a controlled environment in a sitting position and an outdoor test, which involves movement and interaction with the environment. The details of the test will be described in section 9.4 and the results in section 9.8.

9.3.3. Robustness

To prove thesis III (see 4.2) how robust and reliable the system is when one of the measurements falls away, different version of the models have been used. In total, there were three versions of each application. Each of the versions had one or more of the input signals missing. The versions were:

- a version using only physiological data and no context information
- a version where one of the physiological data channels is missing.
- a version with all physiological signals and context information

A comparison between the model without context information and the model with all input channels is done for thesis I (9.9). The results of the comparison between the model with channel loss and the full model is presented in 9.10.

9.4. Test Design

The test was designed to prove the three thesis introduced in chapter 4. For testing, different applications, described in chapter 8 have been modified and used. In the following, test environment, test setting and test setup will be introduced.

9.4.1. Test Environment

The evaluation takes part in two different scenarios. On the one hand indoor tests for a controlled environment. On the other hand outdoor tests, to prove the concept in mobile scenarios that require more attention of the user to the environment.

The tests for the indoor scenario are conducted in the usability lab of Goethe University. Participants were seated on a couch during the test. For the outdoor scenario participants have to walk around the Campus Bockenheim in Frankfurt and through different buildings. In the outdoor scenario participants were guided and given directions by the experimenter. During the outdoor scenario, participants were constantly moving, whereas the participants of the indoor scenario were sitting.

9.4.2. Test groups

The participants are divided into four groups. In the first step, participants are divided in indoor and outdoor group. In each of these two groups, participants are divided further into a non-stress and a stress group as shown in figure 9.4.

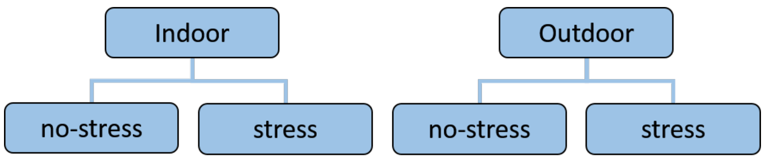


Figure 9.4.: Test groups for evaluation

The stress group gets additional arithmetical tasks during usage of the applications to increase the stress and mental load level. The no-stress group had only to concentrate on the applications. The division into different stress groups was done, to simulate indoor and outdoor situations under different kind of mental load and stress.

9.4.3. Test Procedure

To cover a broader range of applications, two applications with different concept have been chosen. The application VocabTrainer (see 8.3) and the game Zone of Impulse (see 8.1) have been chosen for the study. The vocable trainer application makes primarily use of the cognitive state of an user, where as the game Zone of impulse primarily uses the affective state.

To assess subjective aspects of cognitive and affective state, different questionnaires and surveys have been used and developed. Subjective mental load is measured by the NASA Task Load Index questionnaire, introduced in chapter 2.4.2. Additionally, a questionnaire about user experience was created, addressing several important points for affective and cognitive state, like fun, if the user felt overstrained or if the user felt supported.

The test was divided in several steps. The order of applications was randomized in each test. The sequence of the study is visualized in table 9.1.

Time	Description
5 minutes	Information about the test for the participant
3-5 minutes	Introduction questionnaire and consent form
5-10 minutes	fitting of the sensors and baseline measurement for 5 minutes
	Application 1a
	Short break, NASA-TLX and usability survey
	Application 1b
	Short break, NASA-TLX and usability survey
	Application 1c
	Short break, NASA-TLX and usability survey
	Application 2a
	Short break, NASA-TLX and usability survey
	Application 2b
	Short break, NASA-TLX and usability survey
	Application 2c
	Short break, NASA-TLX and usability survey

Table 9.1.: Test procedure of evaluation

At the beginning, each participant was informed about test procedure and had to fill out a consent form and a demographic questionnaire. After this step, sensors were placed on the participant and a 5 minutes and 20 second long baseline was measured in a sitting position. This part of the test took place in the laboratory, independent of the test group.

In the next step, the main part of the study starts. In randomized order, participants had at first to play three different versions of Zone of Impulse or three different versions of vocable trainer. The three versions of the chosen application itself were also in randomized order. Details on the different versions of the applications will be given in the following subsection. After usage of each version, participants had a short break and to fill out NASA-TLX and the other survey for the current version.

9.4.4. Application Versions

To test different aspects of the model, like context integration or robustness, variations of the original application have been developed. The following versions of the applications ”Zone of Impulse” and ”Vocable Trainer” have been created:

- Zone of Impulse without the model in the background and any adaptation
- Zone of Impulse with full biofeedback model
- Zone of Impulse without EDA

For the partial model only heart rate and context information had been chosen. In physiologically enhanced games usage of EDA is more common instead of heart rate. For the stress test of the model in the background it has been chosen to only use heart rate to test the worst case scenario of channel loss for gaming situations.

In the same way, the application ”vocable trainer” had a version, that was reduced to having only EDA. In a learning application, mental effort based on HRV would be more useful but for the stress test of the model, EDA has been chosen. The versions of ”Vocable Trainer” are:

- Vocable Trainer without the model in the background and any adaptation
- Vocable Trainer with full biofeedback model
- Vocable Trainer without HRV

9.5. Execution of Test and Demographic Survey

The test was executed between June and July 2014. In total 41 people participated in the test. The distribution of participants over the different groups is shown in table 9.2. Four of the participants were female, 36 male with an average age of 25 years. The duration of the test was approximately between 45 to 90 minutes, depending on the test group.

Group	Number of Participants
stress indoor	16
stress outdoor	8
indoor	9
outdoor	8

Table 9.2.: TLX values for Vocable Trainer

Asked about their smartphone experience, participants rated their experience in average with 4.24 (1="not an experienced user" and 6="very experienced user"). In average, participants regularly use the internet on mobile devices rated with 4.38 (1="do not use it", 6="use it very often").

9.6. Impairment by Sensors

After the study, participants were asked if they felt impaired by wearing the sensor during the test. The rating of the sensors is shown in figure 9.5. In total, participants did not feel impaired by wearing the sensors. Average rating over all groups was 1.63, with a standard deviation of 1.03 on a scale from 1="I did not feel impaired" to 6="i did feel impaired". The outdoor scenario involved movement and interaction with the environment like taking stairs, open doors and following directions. In the indoor scenario, participants were sitting on a couch.

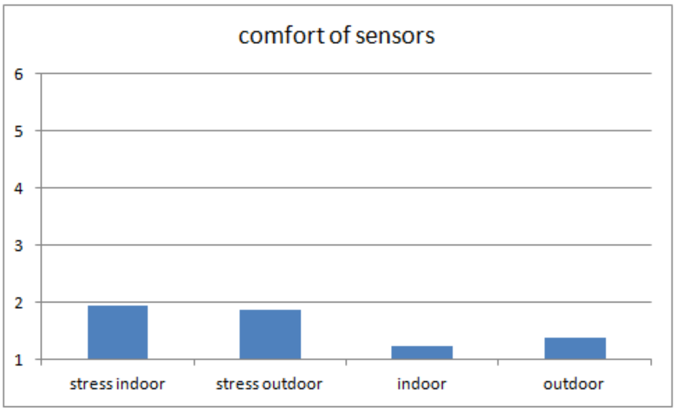


Figure 9.5.: Results of wearing comfort rating (1=did not feel impaired by sensors, 6=did feel impaired by sensors)

The rating shows a small difference between the groups with additional stress and the groups without stress. The best rating was given in the indoor group without stress, followed by the outdoor group with no additional stress. The groups of participants which experienced additional stress, rated the impairment on average with 1.9. Participants without additional stress rated the impairment aspect better, with an average score of 1.3.

In summary, the results show that the chosen sensors are suitable for mobile scenarios and did not impair the participants.

9.7. Comparison of Physiological Versions against Normal Version

In this section, the results of two of the three versions of each of the both applications are compared: the version with the full model and the version without any integration of physiological signals and context information. In the following, the results for vocable trainer and zone of impulse are presented.

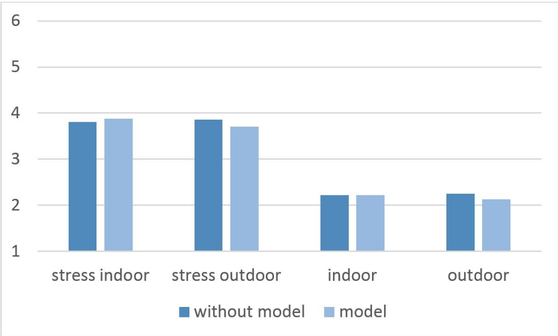
9.7.1. Vocable Trainer

For the vocable trainer, survey results and NASA-TLX were recorded and analyzed. The results show a difference in some points, which will be discussed in the following. The expectation is, that ratings for overstrain are lower or similar in the version with integrated model in contrast to the version without model. In the same way, it is expected, that the ratings for fun and support are the same or slightly higher for the version with the model, depending on the adaptations that were done in the applications. Mental load is expected to be lower in the versions with integrated model.

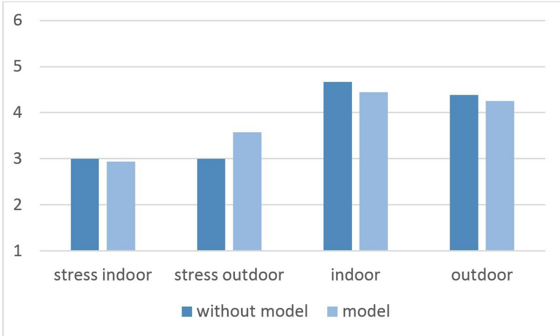
Survey Results

In the survey, three aspects had to be rated: overstrain, support and fun. The results for fun, overstrain and support are shown in figure 9.6 (a), (b) and (c) for the single test groups. Table 9.3 shows the absolute average values for overstrain, support and fun.

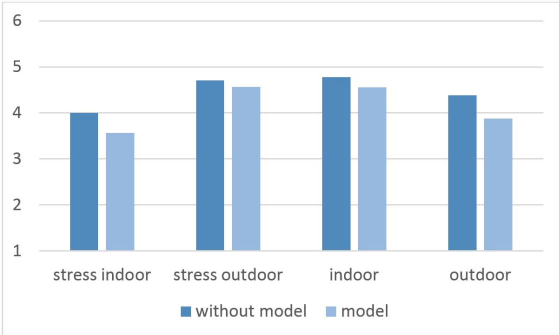
When looking at the average values for the model and the version without model across the test groups, only a difference in the fun ratings exist. Overstrain was rated with 3.2 (no model) and 3.12 (model) on average on a scale from 1 to 6, with 1="i did not feel overstrained" and 6="i felt overstrained". When looking on the aspect of overstrain in the single test groups, in the groups without stress was nearly no difference between the ratings. In the rating of the stress groups, the indoor participants rated the version with no model slightly better, the outdoor participants rated the version with model slightly better. When comparing the group with and the group without model in a Wilcoxon signed rank test, the differences are not significant (p=0.6871).



(a) Overstrain ratings of Vocable Trainer



(b) Support ratings of Vocable Trainer



(c) Fun ratings of Vocable Trainer

Figure 9.6.: TLX rating of vocable trainer for the version without and with model

	No Model	Model
overstrain	3.2	3.12
support	3.66	3.63
fun	4.34	4.02

Table 9.3.: Rating results for the aspects overstrain, support and fun for vocable trainer with and without integrated model

The absolute average values for support, shown in table 9.3 are similar to the results of overstrain and do not significantly differ with values of 3.66 and 3.63 (1="I did not feel supported" and 6="I felt supported"). Looking at the results of the single groups, both version reached nearly the same values and do not differ significantly in the stress indoor and both no stress groups, with a slightly better rating for the version without model. In the stress outdoor group, the version with the physiological model was rated higher, but not on a significant level. A paired Wilcoxon signed rank test between the group with and without model shows no significant difference for the support rating(p=0.8236).

Looking at the average fun rating in table 9.3, there was a difference between the two versions. The version without model reached a slightly better rating with 4.34 in comparison to 4.02 (1="I had no fun", 6="I had fun"). Looking at the results in the single groups shown in figure 9.6(c), the version with no model was always rated better in contrast to the version with model. A Wilcoxon signed rang test shows that fun is rated significantly better in the group without model (p=0.0077). One reason here for might be, that difficulty increases when mental load of an user is low. This increase in difficulty might lead to a decrease in fun.

NASA-TLX results

Besides the survey with aspects of fun, overstrain and support the NASA-TLX was recorded and analyzed for each version of the application. Figure 9.7 shows the results of the rating in the different groups in a graph and table 9.4 the absolute values.

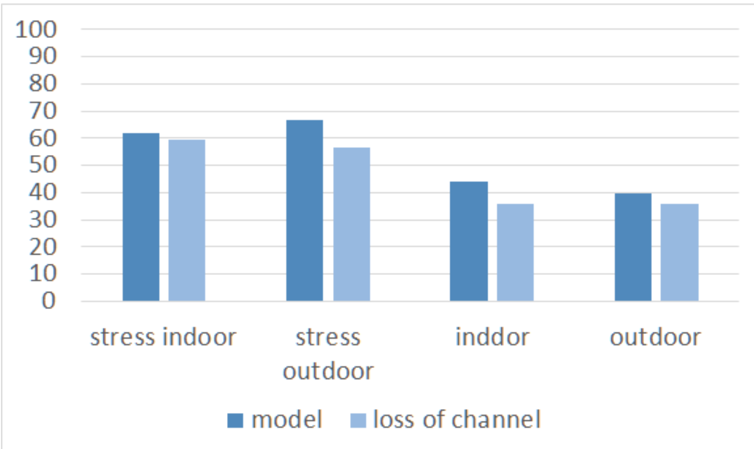


Figure 9.7.: TLX rating of vocable trainer for the version without and with model

	no model	model
stress indoor	61.76	59.2
stress outdoor	66.75	56.75
indoor	44.15	35.89
outdoor	39.46	35.91

Table 9.4.: TLX values the versions with and without engine of vocable trainer

The NASA-TLX produces values between 0 and 100, where 0 stands for a low and 100 for a high mental load. The results show only a small difference between the two stress groups. In both stress groups, participants had lower mental load ratings in the version with integrated model (indoor: 59.2 to 61.76, outdoor: 56.75 to 66.75). In the groups without stress, the differences were higher, especially in the indoor group. In both groups the mental load ratings were lower for the version with integrated model (indoor: 35.89 to 44.15, outdoor: 35.91 to 39.46). A Wilcoxon signed rank test comparing if the version without integrated model has higher mental load values as the version with model leads to the result, that mental load in the version without model is significantly higher (p=0.001408).

9.7.2. Zone of Impulse

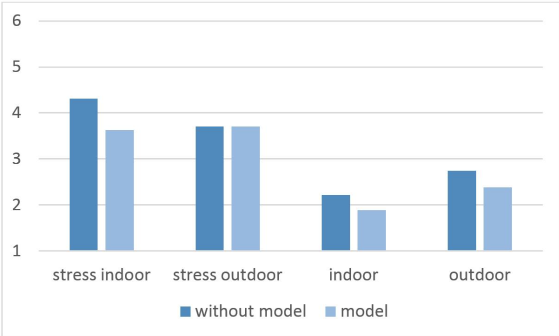
As Zone of Impulse is a game, survey about overstrain, support and fun is more important, as the ratings of NASA-TLX. In the following, results of both are presented, comparing the version without model to a version with the model.

Survey Results

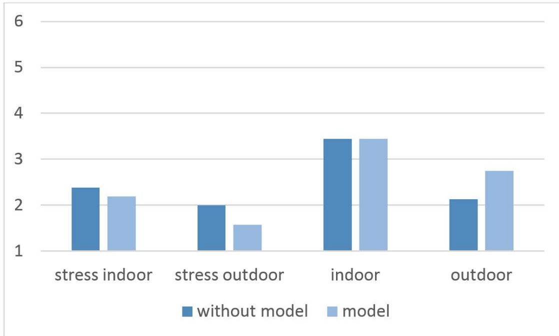
The average rating results for the survey comparing both versions are shown in table 9.5. The results for the single groups are shown in figure 9.8.

The average values show a difference in the overstrain rating and no or only slight difference in fun and support rating. The average value for overstrain in the version with model is 3, in the version without model 3.41 (1="I did not feel overstrained", 6="I felt overstrained"). When looking at the results in the single groups for overstrain, overstrain rating is for the version using the model in every group lower, except the outdoor stress group. As this is the group with the highest amount of stress, participants might not have recognized a difference because of a too high stress load.

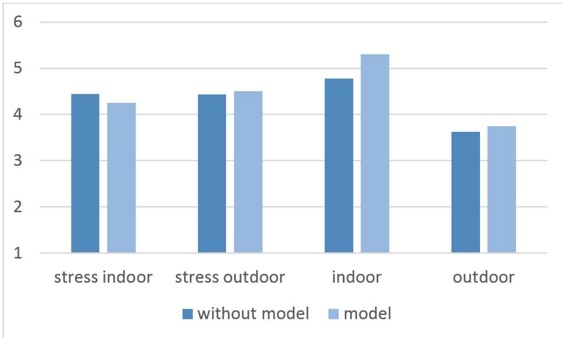
Looking at the average value for support, there was no difference between both versions. When analyzing the results of support rating for the single groups, the result is mixed. Two groups felt supported by the version without the model, one felt supported by the version with model and one group was undecided. The two stress groups felt more supported by the version without model. But in general the rating for the stress groups was very low regarding support, as users seemed to feel not supported in general.



(a) Overstrain rating for Zone of Impulse



(b) Support rating for Zone of Impulse



(c) Fun rating for Zone of Impulse

Figure 9.8.: Rating results for the aspects overstrain, support and fun for Zone of Impulse with and without integrated model

	No Model	Model
overstrain	3.41	3
support	2.46	2.46
fun	4.29	4.37

Table 9.5.: Absolute rating results for the aspects overstrain, support and fun for Zone of Impulse with and without integrated model

The results of the average fun rating were close together, with a rating of 4.29 for the version without model and 4.37 for the version with model (1="I had no fun", 6="I had fun"). When looking at the results of the single groups, the groups are similar to the support rating indifferent about the fun rating. The indoor group without stress clearly preferred the version with integrated model. The outdoor group without stress rated both versions equal. Stress outdoor rated both versions nearly equal and stress indoor rated the version without model only slightly better.

Overall there was no significant difference for support and fun rating between both versions over all groups. Looking at the results of a paired Wilcoxon rank sum test, overstrain was rated significantly better in the group with integrated model with a p-value of 0.02723.

NASA-TLX Results

Besides the already presented aspects, NASA-TLX was also recorded for Zone of Impulse. As it is a game with the main goal fun the expectation was that the results of the NASA-TLX are in general lower, than in comparison to the NASA-TLX ratings of the vocable trainer. The results are visualized in figure 9.9 and the absolute average values of each group are shown in table 9.6.

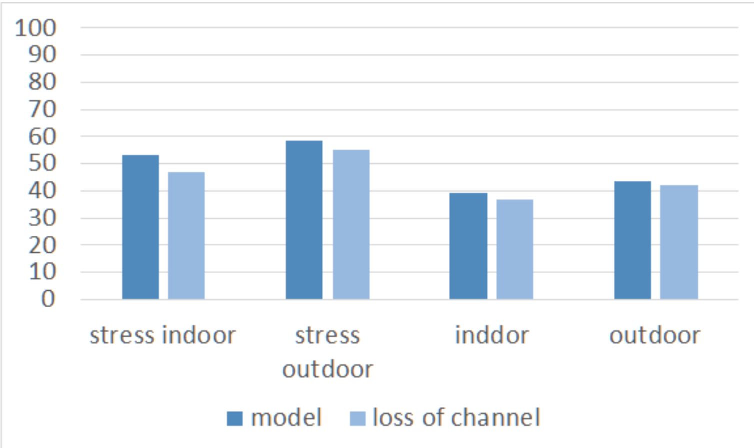


Figure 9.9.: TLX rating of Zone of Impulse for the version without and with model

	no model	model
stress indoor	53.06	46.87
stress outdoor	58.37	55.29
indoor	39.37	36.89
outdoor	43.62	42.21

Table 9.6.: TLX values the versions with and without model of Zone of Impulse

As expected, the average NASA-TLX ratings are lower in contrast to the results of vocable trainer. In the stress groups, there is a significant difference between the ratings of both versions. The version with integrated model got lower ratings for mental load as the version with no integrated model (stress indoor: 53.06 to 46.87, stress outdoor: 58.37 to 55.29).

In the groups without stress, the version with model had also the lower mental load ratings but the difference was smaller (indoor: 39.37 to 36.89, outdoor: 43.62 to 42.21). In this case a Shapiro Wilk test showed, that both data sets are normally distributed. A paired t-test comparing if the version without integrated model has higher mental load values as the version with model leads to the result, that mental load in the version without model is significantly higher (p=0.04861).

9.7.3. Conclusion

Summarizing the results of the comparison between versions with and without model of two different applications, there has to be made a difference between the application that focuses on cognitive state and the game, which focuses on affective state.

In the vocable trainer overstrain and support rating showed no significant difference between both versions. Fun was rated significant better in the version without model. However, the NASA-TLX rating showed a difference in mental load. The version with integrated model leads to significant lower mental load values in contrast to the version without model.

Zone of Impulse had a significant difference in the overstrain rating and no significant difference in the support and fun rating. The version with model had significant better ratings regarding overstrain than the version without. The mental load results showed also significant differences between both versions. The version with integrated model leads in all four groups to lower mental load values than the version without model.

Both applications showed no positive difference in support and fun for the version with model. But both applications showed lower values in mental workload ratings, when using the version with model. Zone of Impulse also showed better rating values for overstrain in the version with model. In conclusion, participants had approximately equal values for support and fun but the model lead to lower workload ratings, as a positive effect of the model.

9.8. Comparison of Indoor and Outdoor Scenarios

To examine if the model works also for outdoor scenarios, the results of indoor and outdoor test groups have been compared. For both, vocable trainer and zone of impulse, the ratings for fun, support and overstrain as well as NASA-TLX are compared for the version with integrated model and are presented in the following.

The expectations are, that the version with model shows no significant worse results in the ratings of outdoor scenario in comparison to the indoor results.

9.8.1. Vocable Trainer

For the vocable trainer results of overstrain, support and fun ratings were compared between indoor and outdoor groups for the version with complete integrated model. The same was done for the results of NASA-TLX rating.

Survey Results

The results for the different ratings of both groups are visualized in figure 9.10. Table 9.7 shows the absolute average values of the single aspects and situations.

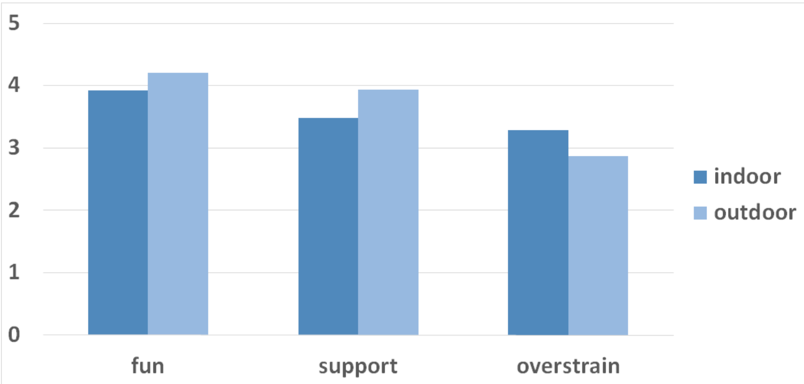


Figure 9.10.: Comparison of ratings of fun, support and overstrain for the indoor and outdoor scenarios of Vocable Trainer with full integrated model

The results showed, that the outdoor group rated the applications slightly better in all three aspects in both versions: fun and support were higher, overstrain was lower. The absolute numbers are presented in table 9.7.

	fun	support	overstrain
indoor	3.92	3.48	3.28
outdoor	4.19	3.88	2.88

Table 9.7.: Absolute ratings of fun, support and overstrain for the indoor and outdoor scenarios of Vocable Trainer with full integrated model

Fun rating has with 0.27 and support with 0.4 difference only a slightly higher rating in outdoor than in the indoor group. With a difference of 0.4 points participants in the outdoor group felt less overstrained. A two-sided Wilcoxon signed rank test comparing the ratings leads to the result that there are no significant differences in the ratings.

	model
stress indoor	59.2
stress outdoor	56.75
indoor	35.89
outdoor	35.46

Table 9.8.: Average values of NASA-TLX rating of vocable trainer

NASA-TLX Results

The results of NASA-TLX rating for the version of vocable trainer with integrated model are presented in figure 9.11. The absolute numbers are shown in table 9.8.

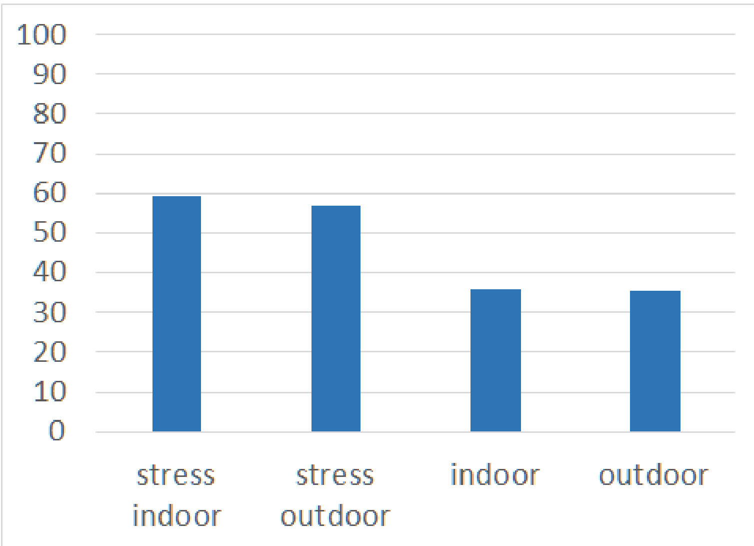


Figure 9.11.: NASA-TLX ratings for Vocable Trainer with integrated model

The figure shows that there is nearly no difference between the values for stress indoor in comparison to stress outdoor, as well as for no stress indoor in comparison to no stress outdoor. Testing the data for indoor and outdoor group with a Shapiro Wilk test for normal distribution leads to the result that both are normally distributed. A two-sided t-test confirms that the differences are not significant.

9.8.2. Zone of Impulse

For Zone of Impulse, the same aspects were analyzed as for the vocable trainer to make a comparison between stress indoor and stress outdoor group, as well as between indoor and outdoor groups without stress.

Survey Results

The results of fun, support and overstrain rating are shown in figure 9.12 as well as the absolute average numbers in table 9.9.

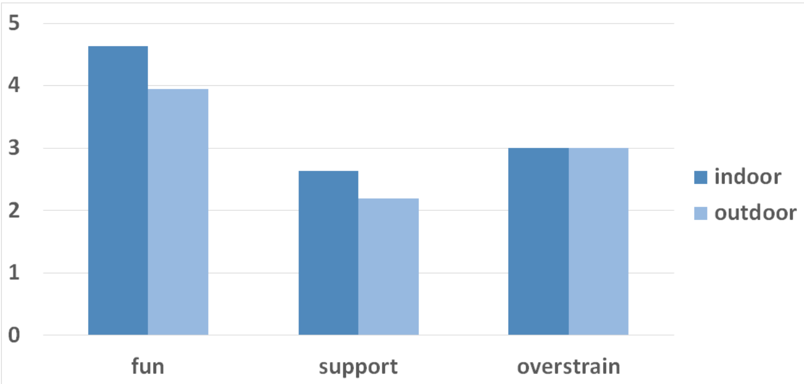


Figure 9.12.: Comparison of ratings of fun, support and overstrain for the indoor and outdoor scenarios of Zone of Impulse

The results showed that the ratings for fun, support and overstrain are slightly lower in the outdoor scenario in comparison to indoor. The absolute average values are shown in table 9.9.

	fun	support	overstrain
indoor	4.64	2.64	3
outdoor	3.95	2.19	3

Table 9.9.: Comparison of ratings of fun, support and overstrain for the indoor and outdoor scenarios of Zone of Impulse

A two-sided Wilcoxon rank sum test shows that there is no significant difference between the indoor and outdoor groups for all three aspects.

NASA-TLX Results

The results of NASA-TLX rating for Zone of Impulse with integrated model are presented in figure 9.13. The absolute numbers are shown in table 9.14.

The figure shows, that there are slightly higher mental load values for outdoor group in comparison to indoor group. The same applies for the comparison between stress indoor

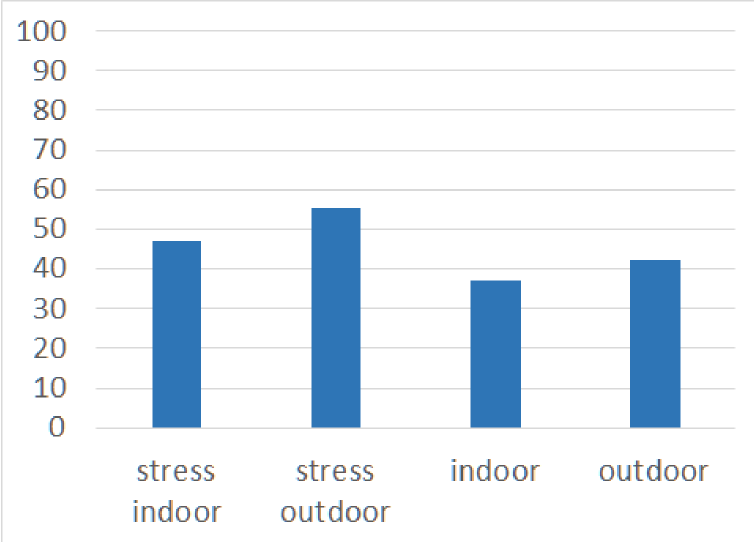


Figure 9.13.: NASA-TLX ratings for Zone of Impulse with integrated model

	model
stress indoor	46.87
stress outdoor	55.29
indoor	36.89
outdoor	42.21

Figure 9.14.: Average values of NASA-TLX rating of Zone of Impulse

and stress outdoor group. Stress outdoor group has slightly higher ratings. Shapiro Wilk tests show that the results of both groups are normally distributed. A two sided t-test shows, that there is no significant difference between indoor and outdoor group as well as between stress indoor and stress outdoor group.

9.8.3. Conclusion

Summarizing the results, vocable trainer showed no significant difference in survey ratings for support, overstrain and fun. Also the rating of NASA-TLX shows no significant difference between the indoor and outdoor group, as well as between the indoor stress and outdoor stress group. The same results were observed for Zone of Impulse.

When looking at the questions raised for thesis II in chapter 4.2, the results indicate to support the question regarding loss of quality in interpretation during mobility, as no significant difference between both situations could be identified in the study. For a test on equivalence, a two one-sided test would be required. For this test, an upper and lower bound for equivalence would be needed to be defined subjectively. For a reliable definition of these bound, more data is needed.

9.9. Context Integration

Another important aspect of this work was the integration of context information. Results of a previous study were used to compare the version with context information of Zone of Impulse with a version, which had the same model but no integrated context information. The test setup was the same, participants played the game for five minutes. Afterwards they had to rate aspects like fun and support. In total, 11 participants took part. As the previous study had only one test group, indoor without additional stress, the results were compared to the indoor group of the actual study.

In both studies, the same ratings scale was used. As the previous study only had fun and support ratings, only these values are compared. Figure 9.15 and table 9.10 show the results of the rating for fun and support for the version of the model with and without integrated context information.

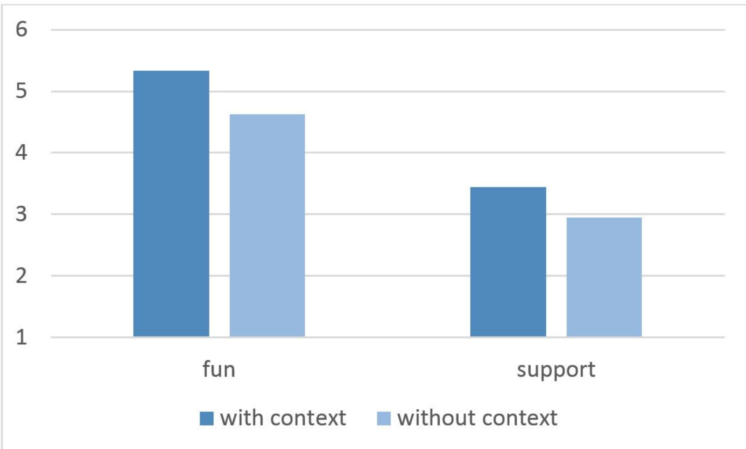


Figure 9.15.: Fun and support rating for version with and without context

	context	no context
fun	5.33	4.64
support	3.44	2.91

Table 9.10.: Fun and support rating for version with and without integrated context information as well as the version without integrated model

The comparison of the results shows, that fun and support were rated higher for the version with integrated context information in contrast to the version without context information. The result of the t-test for support rating leads to a positive statistically tendency that support rating for integrated context is higher, but the result is not statistically significant with a p-value of 0.111. The t-test result for fun rating shows, that fun was rated significantly higher in the version with integrated context, than in the version without integrated context (p=0.04526).

This indicates that the model with integrated context information might lead to better results. A comparison between indoor and outdoor has been done in the previous subchapter 9.8, showing that the version with model leads to no significant different results in indoor and outdoor scenarios.

9.10. Loss of Input Channels

To test the reliability and robustness of the model, a version with one missing input channel has been created for each of the applications. In the following rating results for both applications will be presented.

9.10.1. Vocable Trainer

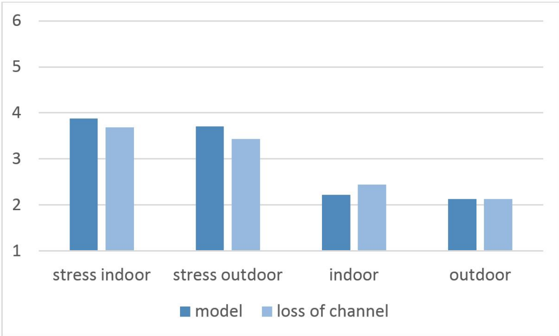
Two versions of the model have been used for this test. An unmodified version with the complete model and one version, which had no heart rate (HR) as input and therefore no HRV. HRV has been chosen as it is expected to be the worst case scenario of channel loss, as HRV is important for cognitive state. In the following results of the survey and NASA-TLX are presented.

Survey Results

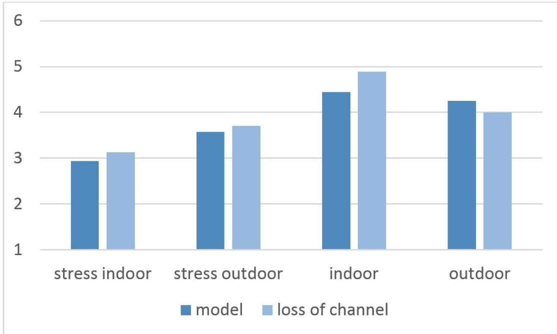
The average results of both versions for overstrain, support and fun are shown in table 9.11. The results for the single test groups are visualized in the figures 9.16.

The average results show, that there are slight differences between the full model and the reduced model. Rating of overstrain, support and fun differ only slightly between the version with model and the version with channel loss. This leads to the assumption that the channel loss can be compensated without a decrease in ratings.

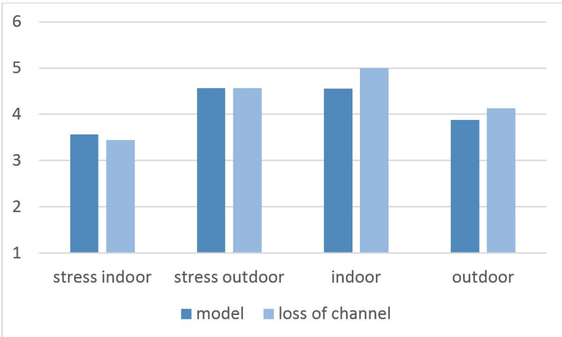
When looking at the results of the stress groups (indoor and outdoor), the values for the reduced model for overstrain was slightly decreased and support increased. The aspect fun was rated nearly the same for stress indoor and outdoor. In the group without stress, the reduced model achieved slightly higher ratings.



(a) Rating for overstrain



(b) Rating for support



(c) Rating for fun

Figure 9.16.: Ratings for overstrain, support and fun for the version with model and with channel loss

	Model	Input loss
overstrain	3.12	3.1
support	3.63	3.8
fun	4.02	4.12

Table 9.11.: Rating results for the aspects overstrain, support and fun for vocable trainer with full model and a version with loss of an input channel

A Wilcoxon signed rank test showed, that the differences in the ratings of support, overstrain and fun were not significant.

NASA-TLX Results

The results of NASA-TLX ratings for mental load are visualized in figure 9.17. Table 9.12 shows the absolute values.

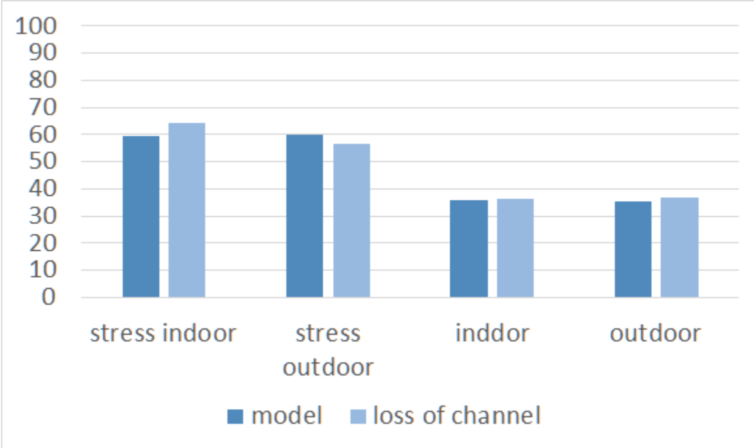


Figure 9.17.: TLX rating of vocable trainer for the version with full and reduced model

The average results shown in table 9.12, imply that there is no significant difference in workload rating in the groups without stress for the mental load rating. In the stress indoor group, mental load was slightly higher for the model with channel loss in contrast to the full model. In the stress outdoor group, the opposite was happening. The model with channel loss was rated slightly better, than the full model.

	model	loss of channel
stress indoor	59.2	65.04
stress outdoor	56.75	57.66
indoor	35.89	36.44
outdoor	35.91	37

Table 9.12.: Average values of NASA-TLX rating of vocable trainer for the version with full and reduced model

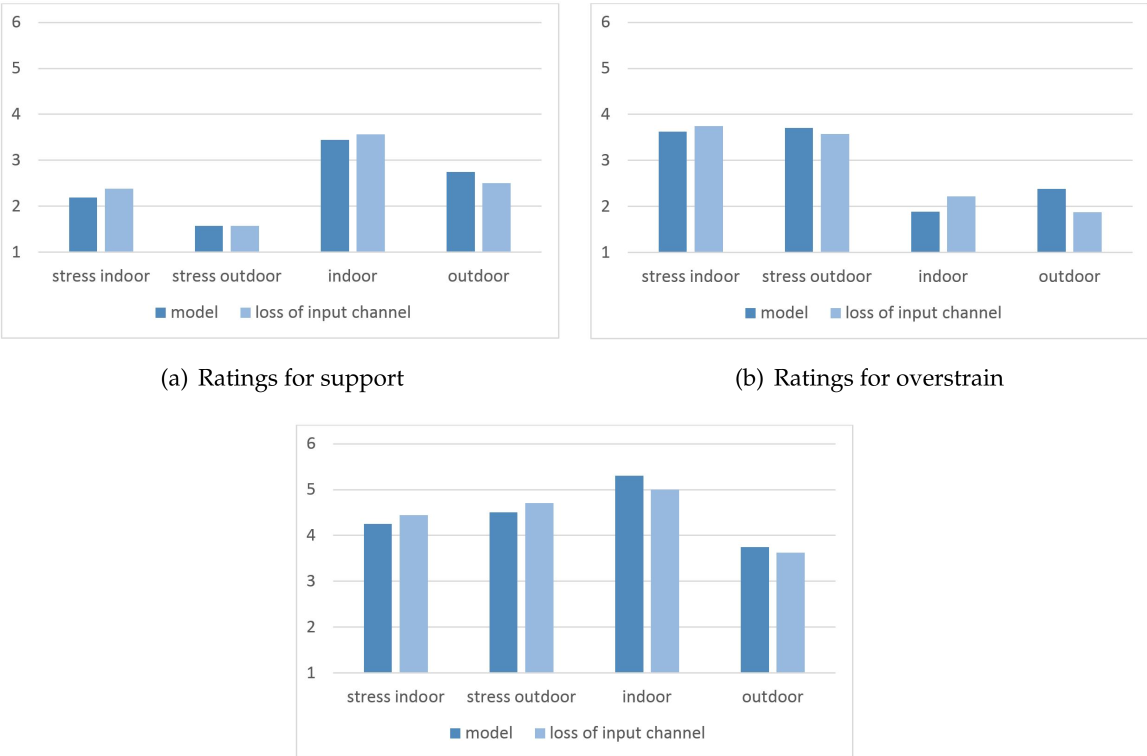
The results of a paired t-test for NASA-TLX ratings show no significant difference between the version with full model and the version with channel loss.

9.10.2. Zone of Impulse

As Zone of Impulse is a game and makes extensive use of the affective state, the version with missing input channel will be reduced by EDA. In the following the results will be presented.

Survey Results

The results for overstrain, support and fun rating for both versions are presented in figure 9.18. The figures show, that there is no big difference between the ratings in the single groups.



(a) Ratings for support

(b) Ratings for overstrain

(c) Ratings for fun

Figure 9.18.: Ratings for support, overstrain and fun for the version with model and a version without a missing input channel

Table 9.13 shows the average ratings for fun, overstrain and support for both versions. The average values show no significant difference between both versions. The difference is less than 0.05 points on the 6 point rating scale for each aspect. When taking a closer look at the results of the single groups, there are no big differences between the most ratings. A Wilcoxon signed rank test showed that there is no significant difference between the two versions for the ratings. This implies, that the reduced model with input channel loss compensates the difference of the lost input channel reliable enough.

	Model	Input loss
overstrain	3	3.05
support	2.46	2.49
fun	4.37	4.37

Table 9.13.: Rating results for the aspects overstrain, support and fun for Zone of Impulse with full model and a version with loss of an input channel

NASA-TLX Results

Looking at the NASA-TLX rating results, the results of the survey can be confirmed. Figure 9.19 shows the results of the single groups, table 9.14 shows the average values.

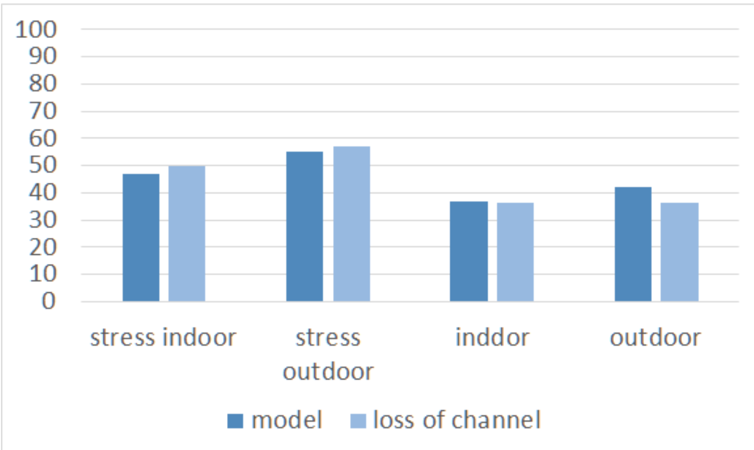


Figure 9.19.: TLX rating of Zone of Impulse for the version without and with model

When taking a closer at the average values, the indoor group as well as the stress outdoor group had nearly no difference between the values. The outdoor group had a difference of 5.67 points between loss of channel and version with full model. The reduced model had in this case the lower mental load rating, implying that mental load was lower than in the full model.

	model	loss of channel
stress indoor	46.87	49.7
stress outdoor	55.29	57.08
indoor	36.89	36.07
outdoor	42.21	36.54

Table 9.14.: Rating results for the aspects overstrain, support and fun for Zone of Impulse with full model and a version with loss of an input channel

A paired t-test shows, that the differences between the full model and the model with channel loss are not significant within the single groups.

9.10.3. Conclusion

The results for this part of the study showed that in both applications, vocable trainer and zone of impulse, the results for the versions with full and reduced model had only small differences in the survey ratings, which turned out to be not significant. In both cases, the NASA-TLX ratings for mental load supported the results. This leads to the assumption, that a channel loss can be compensated.

This study is limited to only a small part of possible situations. Relatively normal situations have been covered, extreme situations might lead to other results. The loss of more than two channels might for example lead to wrong interpretation results as not enough information would be available. In this case the model, as it is used in this study, would have no physiological input if two channels are lost.

9.11. Conclusion

A study was created, testing different versions of two applications under different conditions. In total 41 participants were divided into four different groups: outdoor and indoor as well as an outdoor and indoor group with additional stress. Different configurations of the model have been used: the complete model, a version with loss of an input channel and a version of the application with no integration of the model at all. Summarizing up the results of the study, five different aspects have been presented to support the three theses.

However, the study is limited and covered only a small part of possible aspects that can appear in reality. In the following a conclusion for the three theses will be drawn, based on the conclusion of study results. A comparison of the results to results of related work, as e.g. presented in the chapter 3 is difficult as other aspects were evaluated with different test settings.

9.11.1. Results Thesis I

In our first thesis, we assume that the combination of physiological signals and context information leads to better results in the interpretation of user state. To prove this thesis a model based on a fuzzy logic approach was designed. Different physiological signals and context information have been chosen as input. As output, affective and cognitive state were chosen to cover a broad range of possible applications.

In 9.7 a version without model is compared to a version with full integrated model for two applications. The results showed no significant difference in support and overstrain rating for the application vocable trainer between both versions (with and without model). However, fun was rated better in the version without model, which might be explained by the rising difficulty. Zone of Impulse had a significant difference in the rating of overstrain. Overstrain was rated better in the group with integrated model. Fun and support did not show a significant difference in ratings. Comparing the results of NASA-TLX for both versions, the version with integrated model was rated significantly better for both applications as the mental load values were lower.

In section 9.9 a comparison of a version with the full integrated model and a model without context information, which was used in a previous study was done. The study was only executed for zone of impulse and only part of the rating aspects was covered. The results showed a significant better rating for the aspect fun for the version with full model. Support had no significant difference between both groups, but the calculated p-value for testing if the version with full model is rated better, shows with 0.111 a positive trend.

In conclusion thesis I, the combination of physiological signals and context information, is supported by the results of the comparison in section 9.7. The version with the model combining physiological and context information was rated significantly better than a version without a model for task load rating and partially for overstrain. The thesis is further supported by the significant better fun rating results of the comparison in section 9.9, where the current version with model was compared to a version without context integration.

9.11.2. Results Thesis II

In the second thesis, it was assumed that a general model for mobile scenarios can be created. A general concept has been developed, visualizing the flow between an interpretation controller, the user and the application. The interpretation controller uses the input of physiological signals from the user to estimate the user state. The user state is then given to the application, which decides about possible adaptations. The fuzzy based model is the core of the interpretation controller and was designed to be as flexible as possible, addressing different requirements of mobile scenarios. Additionally, context information influence the interpretation by controlling environmental influencing aspects on the physiological signals as well as an additional information in the interpretation step itself.

In a first step, wearing comfort of sensors was evaluated. In chapter 9.6, the study results are presented asking the user about impairment through sensors during usage. The result showed, that participants did not feel impaired by wearing the sensors in any of the situations during the test.

In a next step presented in 9.8, the results of both applications were compared in indoor and outdoor tests with the full integrated model. The results show no significant difference between usage of the model in indoor and outdoor scenarios for both applications. This leads to the assumption, that the model is also working in outdoor scenarios.

In conclusion thesis II is supported by the results presented in chapter 9.6, which showed that the sensors for measurement did not impair the participants. Furthermore the thesis seems to be supported by the results of the comparison of indoor and outdoor scenario, presented in section 9.8.

9.11.3. Results Thesis III

The third thesis addresses the aspect of robustness and reliability of the model. As the model should be used in mobile scenarios, it is important to address aspects like loss of an input channel or noisy signals. Several mechanisms have been integrated into the interpretation controller, to guarantee a better reliability and robustness.

A comparison between a version with integrated model and a version where one input channel was lost, was presented in section 9.10 for both applications. The results showed that there was not a significant decrease in ratings, when one input channel was missing.

Thesis III, the reliability of the model, is supported by the results presented in section 9.8 and 9.10. The results showed, that a loss of an input channel can be compensated, as well as that the model also seems to work independent of influencing factors from the environment in the outdoor situations of the test. The results might vary for more extreme situations, which were not covered in the study.

10. Conclusion

An insight into the current state of the user, e.g. feelings or mental workload, is a valuable input for applications. Depending on the current state, applications might offer adaptations to increase aspects like usability or performance. Further information not only about the user, but also about the situation the user is in, can be used as additional input. Especially in mobile scenarios, the combination of both can lead to a higher accuracy in interpretation.

The state of the art analysis showed only one example, that combines physiological signals and context information for improvement of user state interpretation. Especially in mobile applications and scenarios, the lack of context might lead to bad or false interpretations, as many physiological signals can be influenced by environmental aspects. The presented state of the art example used activity tracking for control of heart rate, which increases when an user is moving. The results of the state of the art analysis also showed, that there are nearly no examples, which are designed for mobile scenarios. This work tried to fill the gap, by developing a model which addresses mobile scenarios and integration of context information.

In this thesis a model was developed, combining physiological signals and context information as input.As physiological input, EDA and heart rate have been chosen, as they can be measured easily with wireless sensors and offer a broad variety of psychological interpretation.

After an analysis of requirements needed to be fulfilled by a model to address challenges of mobile scenarios, a fuzzy based approach has been chosen to process and interpret the input channels. Based on heart rate variability and context information, the cognitive state of a user is estimated. Electrodermal activity (EDA), heart rate and context information are used to estimate the affective state of a user, based on the affective grid and valence-arousal model.

The implementation of the engine called MUSE (mobile user state estimation), running as a background service on mobile devices, has been presented. The engine was integrated into different presented applications, covering a broad range of application types, like learning of vocabularies, games or an airline information application.

In a first study, the engine has been tested in indoor and outdoor scenarios, with two applications and under different stressful situations. The applications have been tested with different configurations, allowing a first comparison between mobile and nonmobile scenarios as well as a comparison of the model with and without integrated context information. Furthermore the loss of input channels has been evaluated.

In the following sections, the single theses will be discussed and examined and a summary will be given. The chapter closes with a presentation of future work.

10.1. Combining Context and Physiological Input

To prove thesis I, the combination of context information and physiological input, a model was created using both as an input channel. The model creates an output in form of affective and cognitive state of the user. State-of-the-Art analysis showed, that only the work of [SKC$^+$12] combined context information and physiological signals in a first try.

An analysis and categorization of context information has been done. Context information was divided into two categories: context information for controlling influence on physiological signals and context information used for improving interpretation. The first one was used to control influencing aspects on the physiological input, which might lead to false interpretations. The second one was used as additional information enriching the interpretation, e.g. giving information about the location of the user (e.g. at home or work).

After the analysis of possible available information, it was chosen to integrate location, step sensor as well as information from the application itself. The information from the application itself utilizes information about the performance of the user, e.g. gaming score or error rate. The model provides an interface where the application can transmit this information to the interpretation controller.

To answer the question, if context information provided by application and mobile phone can improve the interpretation quality of user state, two aspects have been investigated in the study. On the on hand, the model has been tested by comparing the model with context integration to a version, where only physiological signals and no context information was used. The results showed, that fun was rated better in the version with integrated context information. The result for the aspect support was not significant, but had a slight tendency for better ratings in the version with integrated context.

On the other hand, a comparison between a version with and without integrated model has been done for indoor and outdoor scenarios. The version with model had physiological and context information integrated. The results showed a significant better rating for overstrain for the application zone of impulse and a negative rating for fun in vocable trainer. All other subjective ratings had no significant difference in the rating results. Furthermore the results showed a significant better rating of task load index for the version with integrated model for both applications.

In conclusion, for both evaluated aspects the results were either not significant different or significant better. These results indicate, that thesis I can be answered positively within this work.

10.2. A Fuzzy Logic based Model for Mobile Scenarios

To prove thesis II, a general model has been created, addressing different research questions regarding mobile scenarios. One of the requirements was, that measurement devices for physiological signals should be as small as possible and not impair the user. Therefore, it has been chosen to measure EDA and heart rate (HR), as there are small sensors available, which transmit data wirelessly. For collecting context information, sensors integrated in modern smartphones have been chosen (Global Positioning System (GPS) and step sensor).

As the possible applications that may benefit from the model are from a broad range of different areas, it has been chosen to use affective and cognitive state as output. The interpretation controller processes the incoming information and signals and transmits values for affective and cognitive state to the application, which then decides about usage and adaptation.

The core of the model was based on a fuzzy logic approach, which is used in different examples in current research. After analyzing different methods, the fuzzy logic approach met most of the requirements. In a first step, input signals are fuzzified. Afterwards, the input is transformed to affective and cognitive state in two steps. In the first step valence, arousal and mental load are determined, based on a fuzzy rule set and the fuzzified input. Afterwards these values and additional context information are used to calculate the values for affective and cognitive state, based on a second fuzzy rule set.

The output of cognitive state can have one of four different states: low, medium, high and very high. These values express how cognitive occupied an user is. The affective state on the other hand is divided into eight states, which can each have one of four different values. The affective state interpretation is based on the circumplex model and affective grid of Russell, which are based on valence and arousal. The eight states have been defined based on the affects on the circumplex. Each of this eight affective states can have one of the following four values: very low, low, medium and high.

Studies were done to prove the correct interpretation by the model. The results of cognitive state seem to correlate with the results of subjective ratings in the presented studies. As number of participants in studies were low, further studies need to be done for proving significance of the result.

For thesis II, the following questions have been raised in chapter 4.2:

- Can a model be defined that supports different types of applications, e.g. learning and entertainment?

- Can a model be defined, that supports the usage in mobile scenarios, e.g. usage of an application during travel without loss of quality in interpretation and without impairing the user?

To answer the first question, the model was evaluated in several studies with applications making use of cognitive state as well as applications making use of the affective state. In the study the model was evaluated with a game and a vocable trainer. For the different aspects, both showed compareable results.

To answer the second question, a comparison between usage of the applications in indoor and outdoor scenarios was done. The results showed no significant difference between the usage of the model in indoor and outdoor sceanrios.

Additionally, a study about impairment of the users through sensors was done, to answer the question if a model can be defined, that meets the criteria to not impair the user in mobile scenarios. The results showed, that participants did not feel impaired by the sensors in outdoor and indoor scenarios. Finally, indoor and outdoor scenarios were compared, to prove that the model is suitable for mobile situations. The results showed, that the version with model performed also better in this comparison than the version without model.

10.3. Reliability and Robustness

For proving thesis III, the reliability and robustness of the model, different mechanisms have been integrated in the model to ensure these aspects. The first step in the model, to ensure robustness, is the signal check. In this step, availability of all signals is checked as well as a possible corruptness of the signals. In a second step, reliability and robustness is increased by context information to control influencing factors on physiological signals. This step is done in the signal processing, before signals get fuzzified. A further method is to reduce the number of output states, leading to less details on affective and cognitive state, but more reliable results.

If an input channel is lost or corrupted, other channels can be used to compensate the loss. For example, loss of heart rate, which is used for valence might be compensated by context information like performance, assuming that a high performance might more likely be correlated to a positive valence than the other way round.

Depending on which channel is lost, estimation quality might vary. As the context information about performance provided by the application is only optional, a loss of the movement channel might influence the accuracy of estimation dramatically. As heart rate, which is used for valence estimation, is easily influenced by movement and movement cannot be recognized, valence cannot be estimated if performance is not available.

When losing heart rate as input channel, performance value is needed as well to compensate the valence value and the mental load estimation for the cognitive state. Cognitive state estimation based on performance will only give a tendency if the user is in a medium state or not. If the user is not in a medium state, the performance could be low because of a very low mental load or because of a high mental load. Aspects of EDA might be used in this case, to distinguish between very low and very high cognitive state, but has not been evaluated yet.

A reduction of output states reduces the eight affective states to only four. This might still be enough for most applications and produces more reliable results, depending on which input channel got lost. Further possible but not yet implemented functionalities for improvement of reliability and robustness will be discussed in section 10.5.

The thesis was validated in the test by a combination of different aspects that were tested. A comparison of tests in indoor and outdoor situation showed that influencing factors seem to be controlled. When comparing the full version of the model to a version where an input channel is lost, the results showed that this could be compensated. The ratings showed no significant difference within two tested applications for different situations.

In conclustion, the results indicate that thesis III can be validated within this work, based on the results of channel loss evaluation within this work. However, the study was limited to only a few possible situations out of the situations that could happen in everydays life.

10.4. Summary

Summarizing the previously in detail discussed results of this thesis, two improvements are achieved by the presented model. On the one hand improvements in the interpretation of physiological signals by adding context information have been achieved. On the other hand, the developed model supports the usage of physiological signals in mobile applications by controlling influencing aspects and giving a reliable estimation of user state even when one input channel gets lost. The study only evaluated a small part of possible applications types and scenarios, but the results showed a benefit of the model for the covered applications and scenarios.

The state-of-the-art analysis showed, that models for estimating user state based on physiological signals lack from the integration of additional context information. The presented models concentrated nearly all only on physiological signals for user state estimation. As physiological signals can be influenced by different environmental aspects, these models may not work appropriate under non-laboratory conditions. The presented context-sensitive interfaces on the other hand do not give an insight in to the state of the user. With the model presented in this work, context-information has been integrated and used to overcome the challenge of controlling influencing factors as well as improving user state interpretation by environmental aspects.

The combination of physiological signals and context information in the model also led to the suitability of the engine for mobile applications. Presented state-of-the-art work was designed for stationary situations not addressing the aspects of mobile scenarios. The developed model was integrated into different mobile applications. Studies showed a benefit of the application versions with integrated engine in comparison to the versions without integration, as well as similar results comparing controlled indoor with outdoor scenarios.

10.5. Future Work

With the proposed concept and model, it has been proven that physiological signals and context information can be combined for the interpretation of user state in mobile scenarios. But this work also showed that there are still aspects which can be examined further and raise additional questions.

The model was tested and designed with EDA and HR as input. As mobile technologies are evolving fast, other sensors might become available, offering different kinds of physiological signals as an input. The model was designed to be extendable. For an extension by other sensors, different aspects have to be considered and examined. New sensors might require new kind of signal preprocessing and context information to control influencing factors.

Other possible future additions to the model are new context information channels. Only few context information has been integrated to the model, many more exist and can be measured by integrated smartphone sensors. Besides that, new devices like e.g. google glass or smartwatches offer new context information. Connected homes might also add valuable information about everyday situations, which can improve interpretation.

The applications the model was tested with were mainly applications developed for smartphones. In a further step, the information of user state interpretation could also be offered to other devices or services. For example, the information about user state could be combined in an office with a sign at the door. Based on the user state, the sign at the door could visualize the current state and a person might be disturbed or not. When offering the data to other services or applications, privacy issues arise that need to be addressed. In this work, the interpreted data never left the local smartphone, but for future applications privacy issues need to be examined and discussed.

In the studies that were presented in this work not all aspects could be covered. Further studies should be done, giving an estimation for the loss of reliability depending on which input channels are available or lost. An interesting point would be to give input channels a weighting, e.g. rating EDA higher for the determination of arousal than other measures. This could lead to more reliable interpretations if one physiological signals delivers data that is not corrupted but contradictory to other information.

The model has only been tested primary with two applications. Tests and studies with other applications are needed to verify the model more in detail. Besides that, the model was tested in indoor and outdoor scenarios, but extended tests over a bigger time span would be useful as the tests only took 90 minutes at maximum. In a longer time span other effects might appear that might not have been considered so far.

This work was a first step in the direction of using physiological data as input beyond controlled environments. Several other aspects arise from this work, which could be examined in further future work.

Bibliography

[AD99] Gregory D Abowd and Anind K Dey. Towards a better understanding of context and context-awareness. In *Handheld and Ubiquitous Computing*, page 304. Springer, 1999.

[Aff11] Affectiva. Affectiva q sensor 2.0 protocol. Brochure, 2011.

[APS+14] Daniel Afergan, Evan M Peck, Erin T Solovey, Andrew Jenkins, Samuel W Hincks, Eli T Brown, Remco Chang, and Robert JK Jacob. Dynamic difficulty using brain metrics of workload. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*, pages 3797–3806. ACM, 2014.

[BBC16] BBC. Calculating target zones and thresholds of training, July 2016.

[BDR07] Matthias Baldauf, Schahram Dustdar, and Florian Rosenberg. A survey on context-aware systems. *International Journal of Ad Hoc and Ubiquitous Computing*, 2(4):263–277, 2007.

[BL94] Margaret M Bradley and Peter J Lang. Measuring emotion: the self-assessment manikin and the semantic differential. *Journal of behavior therapy and experimental psychiatry*, 25(1):49–59, 1994.

[BL00] Margaret M Bradley and Peter J Lang. Measuring emotion: Behavior, feeling, and physiology. *Cognitive neuroscience of emotion*, 25:49–59, 2000.

[BMEL08] Margaret M Bradley, Laura Miccoli, Miguel A Escrig, and Peter J Lang. The pupil as a measure of emotional arousal and autonomic activation. *Psychophysiology*, 45(4):602–607, 2008.

[Bou92] Wolfram Boucsein. *Electrodermal Activity*. Springer, 1992.

[CAF13] Pablo Cingolani and Jesús Alcalá-Fdez. jfuzzylogic: a java library to design fuzzy logic controllers according to the standard for fuzzy control programming. pages 61–75, 2013.

[CCW03] Kirsten Cater, Alan Chalmers, and Greg Ward. Detail to attention: exploiting visual tasks for selective rendering. In *ACM International Conference Proceeding Series*, volume 44, pages 270–280, 2003.

[Cor15a] Google Corporation. The connected consumer survey 2014 / 2015, february 2015.

[Cor15b] Microsoft Corporation. Meet cortana, february 2015.

[Cox92] Earl Cox. Fuzzy fundamentals. *Spectrum, IEEE*, 29(10):58–61, 1992.

[CR02] Neil A. Campbell and Jane B. Reece. *Biology*. Pearson Education, 2002.

[Csi91] Mihaly Csikszentmihalyi. *Flow: The psychology of optimal experience*, volume 41. 1991.

[CTB07] John T Cacioppo, Louis G Tassinary, and Gary Berntson. *Handbook of psychophysiology*. Cambridge University Press, 2007.

[CV04] Daniel Chen and Roel Vertegaal. Using mental load for managing interruptions in physiologically attentive user interfaces. In *CHI'04 extended abstracts on Human factors in computing systems*, pages 1513–1516. ACM, 2004.

[DC07] Andrew Dekker and Erik Champion. Please biofeed the zombies: enhancing the gameplay and display of a horror game using biofeedback. *Proc. of DiGRA*, pages 550–558, 2007.

[DCP+07] David M Diamond, Adam M Campbell, Collin R Park, Joshua Halonen, and Phillip R Zoladz. The temporal dynamics model of emotional memory processing: a synthesis on the neurobiological basis of stress-induced amnesia, flashbulb and traumatic memories, and the yerkes-dodson law. *Neural plasticity*, 2007, 2007.

[Dey01] Anind K. Dey. Understanding and using context. *Personal Ubiquitous Comput.*, 5(1):4–7, January 2001.

[DNYP10] Anders Drachen, Lennart E. Nacke, Georgios Yannakakis, and Anja Lee Pedersen. Correlation between heart rate, electrodermal activity and player experience in first-person shooter games. In *Proceedings of the 5th ACM SIGGRAPH Symposium on Video Games*, Sandbox '10, pages 49–54, New York, NY, USA, 2010. ACM.

[Ede72] Robert Edelberg. Electrical activity of the skin: Its measurement and uses in psychophysiology. *Handbook of psychophysiology*, 12:1011, 1972.

[Ein95] Willem Einthoven. Ueber die form des menschlichen electrocardiogramms. *Pflügers Archiv European Journal of Physiology*, 60(3):101–123, 1895.

[Ele13] Polar Electro. Polar h6 heart rate sensor accessory manual english. Broschure, last called on 27.11.2014, 2013.

[EVM94] Gary D Ellis, Judith E Voelkl, and Catherine Morris. Measurement and analysis issues with explanation of variance in daily experience using the flow model. *Journal of Leisure Research*, 1994.

[Fai09] Stephen H Fairclough. Fundamentals of physiological computing. *Interacting with computers*, 21(1):133–145, 2009.

[Fri86] Nico H Frijda. *The emotions*. Cambridge University Press, 1986.

[GS05] Alex Graves and Jürgen Schmidhuber. Framewise phoneme classification with bidirectional lstm and other neural network architectures. *Neural Networks*, 18(5):602–610, 2005.

[Har06] Sandra G Hart. Nasa-task load index (nasa-tlx); 20 years later. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 50, pages 904–908. Sage Publications, 2006.

[Heb55] Donald Olding Hebb. Drives and the cns (conceptual nervous system). *Psychological review*, 62(4):243, 1955.

[HP05] Jennifer A Healey and Rosalind W Picard. Detecting stress during real-world driving tasks using physiological sensors. *Intelligent Transportation Systems, IEEE Transactions on*, 6(2):156–166, 2005.

[HS88] Sandra G Hart and Lowell E Staveland. Development of nasa-tlx (task load index): Results of empirical and theoretical research. *Advances in psychology*, 52:139–183, 1988.

[HTS02] Gabriela Husain, William Forde Thompson, and E Glenn Schellenberg. Effects of musical tempo and mode on arousal, mood, and spatial abilities. *Music Perception*, 20(2):151–171, 2002.

[IDC16] IDC. Smartphone os market share, 2015 q2, march 2016.

[Inc15a] Apple Inc. Siri. dein wunsch ist ihm befehl., february 2015.

[Inc15b] Google Inc. Google now ” die richtigen informationen zur richtigen zeit, february 2015.

[IPDK08] Wijnand IJsselsteijn, Karolien Poels, and YAW De Kort. The game experience questionnaire: Development of a self-report measure to assess player experiences of digital games. *TU Eindhoven, Eindhoven, The Netherlands*, 2008.

[IS03] Jadwiga Indulska and Peter Sutton. Location management in pervasive systems. In *Proceedings of the Australasian information security workshop conference on ACSW frontiers 2003-Volume 21*, pages 143–151. Australian Computer Society, Inc., 2003.

[Jex88] Henry R Jex. Measuring mental workload: Problems, progress, and promises. *Advances in Psychology*, 52:5–39, 1988.

[JM96] Susan A Jackson and Herbert W Marsh. Development and validation of a scale to measure optimal experience: The flow state scale. *Journal of sport and exercise psychology*, 18:17–35, 1996.

[KJK81] Paul R Kleinginna Jr and Anne M Kleinginna. A categorized list of emotion definitions, with suggestions for a consensual definition. *Motivation and emotion*, 5(4):345–379, 1981.

[Kle13] Alex Klein. Adaptiver vokabel trainer auf basis physiologischer daten. Bachelor's thesis, University of Frankfurt, Germany, October 2013.

[Kol01] Gina Kolata. Maximum heart rate theory is challenged. *New York Times. http://query. nytimes. com/gst/fullpage. html*, 2001.

[Kol14] Christian Kollmeier. Darstellung und adaption mobiler user interfaces auf basis physiologischer daten. Master's thesis, University of Frankfurt, Germany, February 2014.

[Kra91] Arthur F Kramer. Physiological metrics of mental workload: A review of recent progress. *Multiple-Task performance*, pages 279–328, 1991.

[LG01] Ameersing Luximon and Ravindra S Goonetilleke. Simplified subjective workload assessment technique. *Ergonomics*, 44(3):229–243, 2001.

[LHMP13] Iolanda Leite, Rui Henriques, Carlos Martinho, and Ana Paiva. Sensors in the wild: Exploring electrodermal activity in child-robot interaction. In *Proceedings of the 8th ACM/IEEE International Conference on Human-robot Interaction*, HRI '13, pages 41–48, Piscataway, NJ, USA, 2013. IEEE Press.

[Lin91] Wolfgang Linden. What do arithmetic stress tests measure? protocol variations and cardiovascular responses. *Psychophysiology*, 28(1):91–102, 1991.

[LML$^+$10] Nicholas D Lane, Emiliano Miluzzo, Hong Lu, Daniel Peebles, Tanzeem Choudhury, and Andrew T Campbell. A survey of mobile phone sensing. *Communications Magazine, IEEE*, 48(9):140–150, 2010.

[LRS05] Changchun Liu, Pramila Rani, and Nilanjan Sarkar. An empirical study of machine learning techniques for affect recognition in human-robot interaction. In *Intelligent Robots and Systems, 2005.(IROS 2005). 2005 IEEE/RSJ International Conference on*, pages 2662–2667. IEEE, 2005.

[LV71] David T Lykken and Peter H Venables. Direct measurement of skin conductance: A proposal for standardization. *Psychophysiology*, 8(5):656–672, 1971.

[MA07] Regan L. Mandryk and M. Stella Atkins. A fuzzy physiological approach for continuously modeling emotion during interaction with play technologies. *International Journal of Human-Computer Studies*, 65(4):329 – 347, 2007. ¡ce:title¿Evaluating affective interactions¡/ce:title¿.

[MAI06] Regan L. Mandryk, M. Stella Atkins, and Kori M. Inkpen. A continuous and objective evaluation of emotional experience with interactive play environments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '06, pages 1027–1036, New York, NY, USA, 2006. ACM.

[Man08] Regan L Mandryk. Physiological measures for game evaluation. *Game usability: Advice from the experts for advancing the player experience*, pages 207–235, 2008.

[MBC+96] Marek Malik, J Thomas Bigger, A John Camm, Robert E Kleiger, Alberto Malliani, Arthur J Moss, and Peter J Schwartz. Heart rate variability standards of measurement, physiological interpretation, and clinical use. *European heart journal*, 17(3):354–381, 1996.

[Mes88] N Meshkati. Heart rate variability and mental workload assessment. *Advances in Psychology*, 52:101–115, 1988.

[Mor79] Neville Moray. *Mental workload: Its theory and measurement*, volume 8. Plenum Publishing Corporation, 1979.

[MP43] Warren S McCulloch and Walter Pitts. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5(4):115–133, 1943.

[NC02] Jeanne Nakamura and Mihaly Csikszentmihalyi. The concept of flow. *Handbook of positive psychology*, pages 89–105, 2002.

[NDJ+09] David Nunan, Gay Donovan, DJORDJE G Jakovljevic, Lynette D Hodges, GR Sandercock, and DAVID A Brodie. Validity and reliability of short-term heart-rate variability from the polar s810. *Medicine and science in sports and exercise*, 41(1):243–250, 2009.

[NGL10] Lennart E Nacke, Mark N Grimshaw, and Craig A Lindley. More than a feeling: Measurement of sonic user experience and psychophysiology in a first-person shooter game. *Interacting with Computers*, 22(5):336–343, 2010.

[NGP11] Mihalis A Nicolaou, Hatice Gunes, and Maja Pantic. Continuous prediction of spontaneous affect from multiple cues and modalities in valence-arousal space. *Affective Computing, IEEE Transactions on*, 2(2):92–105, 2011.

[NKLM11] Lennart Erik Nacke, Michael Kalyn, Calvin Lough, and Regan Lee Mandryk. Biofeedback game design: using direct and indirect physiological control to enhance game interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, pages 103–112, New York, NY, USA, 2011. ACM.

[NL08] Lennart Nacke and Craig A Lindley. Flow and immersion in first-person shooters: measuring the player's gameplay experience. In *Proceedings of the 2008 Conference on Future Play: Research, Play, Share*, pages 81–88. ACM, 2008.

[NL10] Lennart E Nacke and Craig A Lindley. Affective ludology, flow and immersion in a first-person shooter: Measurement of player experience. *arXiv preprint arXiv:1004.0248*, 2010.

[NMM12] Domen Novak, Matjaž Mihelj, and Marko Munih. A survey of methods for data fusion and system adaptation using autonomic nervous system responses in physiological computing. *Interacting with computers*, 24(3):154–172, 2012.

[NRON13] Pedro Alves Nogueira, Rui Rodrigues, Eugénio C Oliveira, and Lennart E Nacke. Guided emotional state regulation: Understanding and shaping players' affective experiences in digital games. In *AIIDE*, 2013.

[oH13] National Institutes of Health. Pulse. *Medline Plus, US National Library of Medicine. http://www.nlm.nih.gov/medlineplus/ency/article/003399.htm , accessed November 21st 2014*, 2013.

[OTRK05] Antti Oulasvirta, Sakari Tamminen, Virpi Roto, and Jaana Kuorelahti. Interaction in 4-second bursts: The fragmented nature of attentional resources in mobile hci. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '05, pages 919–928, New York, NY, USA, 2005. ACM.

[PBB95] Alan T Pope, Edward H Bogart, and Debbie S Bartolome. Biocybernetic system evaluates indices of operator engagement in automated task. *Biological psychology*, 40(1):187–195, 1995.

[Pic95] Rosalind Wright Picard. Affective computing. 1995.

[PKSH10] Oskar Palinko, Andrew L Kun, Alexander Shyrokov, and Peter Heeman. Estimating cognitive load using remote eye tracking in a driving simulator. In *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications*, pages 141–144. ACM, 2010.

[PSP10] Ming-Zher Poh, Nicholas C Swenson, and Rosalind W Picard. A wearable sensor for unobtrusive, long-term assessment of electrodermal activity. *Biomedical Engineering, IEEE Transactions on*, 57(5):1243–1252, 2010.

[PSS+14] Corinna Peifer, André Schulz, Hartmut Schächinger, Nicola Baumann, and Conny H Antoni. The relation of flow-experience and physiological arousal under stress - can u shape it? *Journal of Experimental Social Psychology*, 53:62–69, 2014.

[PTS+10] Arttu Perttula, Pauliina Tuomi, Marko Suominen, Antti Koivisto, and Jari Multisilta. Users as sensors: Creating shared experiences in co-creational spaces by collective heart rate. In *Proceedings of the 14th International Academic MindTrek Conference: Envisioning Future Media Environments*, MindTrek '10, pages 41–48, New York, NY, USA, 2010. ACM.

[PVH01] Rosalind W. Picard, Elias Vyzas, and Jennifer Healey. Toward machine emotional intelligence: Analysis of affective physiological state. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 23(10):1175–1191, 2001.

[RC12] Lisa Rademakers and Daniel Coleman. *Spinoff, 2011: NASA Technologies Benefit Society*. Government Printing Office, 2012.

[RDMP04] Susana Rubio, Eva Díaz, Jesús Martín, and José M Puente. Evaluation of subjective mental workload: A comparison of swat, nasa-tlx, and workload profile methods. *Applied Psychology*, 53(1):61–86, 2004.

[Reh13] Arthur Rehm. Adaptives management von mobilen anwendungen und services auf basis von physiologischen daten. Bachelor's thesis, University of Frankfurt, Germany, August 2013.

[Rei11] Katharina Reitz. Smartphone games based on physiological data, 2011.

[RN88] Gary B Reid and Thomas E Nygren. The subjective workload assessment technique: A scaling procedure for measuring mental workload. *Advances in psychology*, 52:185–218, 1988.

[RN04] Stuart Russell and Peter Norvig. Künstliche intelligenz. *Ein moderner Ansatz, München*, 2004.

[Ros09] Timothy J Ross. *Fuzzy logic with engineering applications*. John Wiley & Sons, 2009.

[RPM98] Nick S Ryan, Jason Pascoe, and David R Morse. Enhanced reality fieldwork: the context-aware archaeological assistant. In *Computer applications in archaeology*. Tempus Reparatum, 1998.

[RSI98] Dennis W Rowe, John Sibert, and Don Irwin. Heart rate variability: Indicator of user state as an aid to human-computer interaction. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 480–487. ACM Press/Addison-Wesley Publishing Co., 1998.

[RSK12] Katharina Reitz, Claudia Stockhausen, and Detlef Krömker. Zone of impulse: physiological data enhanced gaming. In *Proceedings of the 14th international conference on Human-computer interaction with mobile devices and services companion*, MobileHCI '12, pages 221–224, New York, NY, USA, 2012. ACM.

[RSSA03] Pramila Rani, Nilanjan Sarkar, Craig A Smith, and Julie A Adams. Affective communication for implicit human-machine interaction. In *Systems, Man and Cybernetics, 2003. IEEE International Conference on*, volume 5, pages 4896–4903. IEEE, 2003.

[Rus80] James A. Russell. A circumplex model of affect. *Journal of Personality and Social Psychology*, 39(6):1161 – 1178, 1980.

[RWM89] James A Russell, Anna Weiss, and Gerald A Mendelsohn. Affect grid: a single-item scale of pleasure and arousal. *Journal of personality and social psychology*, 57(3):493, 1989.

[SBVL11] Christian Seeger, Alejandro Buchmann, and Kristof Van Laerhoven. myhealthassistant: A phone-based body sensor network that captures the wearer's exercises throughout the day. In *Proceedings of the 6th International Conference on Body Area Networks*, BodyNets '11, pages 1–7, ICST, Brussels, Belgium, Belgium, 2011. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering).

[She08] Lauralee Sherwood. *Human physiology: from cells to systems*. Cengage Learning, 2008.

[SIG11] Bluetooth SIG. Bluetooth profile specification - heart rate profile. Brochure, 2011.

[SK11] Claudia Stockhausen and Detlef Krömker. Measuring mental load with a polar heart rate monitor. In *R. Dörner, D. Krömker (Eds.), Proceedings of the first ITG/GI Workshop on Self Integrating Systems for Better Living Environments .*, pages 45–49. Shaker Verlag, 2011.

[SKC+12] Feng-Tso Sun, Cynthia Kuo, Heng-Tze Cheng, Senaka Buthpitiya, Patricia Collins, and Martin Griss. Activity-aware mental stress detection using physiological sensors. In *Mobile Computing, Applications, and Services*, pages 211–230. Springer, 2012.

[SMFC87] Stephanie A Shields, Kathleen A MacDowell, Sarah B Fairchild, and Mary L Campbell. Is mediation of sweating cholinergic, adrenergic, or both? a comment on the literature. *Psychophysiology*, 24(3):312–319, 1987.

[Smy12] Justine Smyzek. Ein herzratenbasiertes smartphonespiel, 2012.

[SRQ01] Robert Morris Stern, William J Ray, and Karen S Quigley. *Psychophysiological recording*. Oxford University Press, 2001.

[SSK11] Claudia Stockhausen, Daniel Schiffner, and Detlef Krömker. Adaptation of user interfaces based on physiological data. In *S. Schmid, M. Elepfandt, J. Adenauer, A. Lichtenstein (Eds.), 9. Berliner Werkstatt Mensch-Maschine-Systeme*, pages 78–79. VDI, 2011.

[ST94] Bill N Schilit and Marvin M Theimer. Disseminating active map information to mobile hosts. *Network, IEEE*, 8(5):22–32, 1994.

[STC04] John Shawe-Taylor and Nello Cristianini. *Kernel methods for pattern analysis*. Cambridge university press, 2004.

[Str35] J Ridley Stroop. Studies of interference in serial verbal reactions. *Journal of experimental psychology*, 18(6):643, 1935.

[SW15] Robert Sedgewick and Kevin Wayne. Algorithms, 4th edition, java algorithms and clients, february 2015.

[TBBM86] JN Tsanakas, OM Bannister, AW Boon, and RD Milner. The'sport-tester': a device for monitoring the free running test. *Archives of disease in childhood*, 61(9):912, 1986.

[Tec10] Zephyr Techonology. Zephyr hxm bt. Brochure, 2010.

[TMS01] Hirofumi Tanaka, Kevin D Monahan, and Douglas R Seals. Age-predicted maximal heart rate revisited. *Journal of the American College of Cardiology*, 37(1):153–156, 2001.

[TNL+14] Mika P Tarvainen, Juha-Pekka Niskanen, Jukka A Lipponen, Perttu O Ranta-Aho, and Pasi A Karjalainen. Kubios hrv–heart rate variability analysis software. *Computer methods and programs in biomedicine*, 113(1):210–220, 2014.

[TOTK04] Sakari Tamminen, Antti Oulasvirta, Kalle Toiskallio, and Anu Kankainen. Understanding mobile contexts. *Personal and ubiquitous computing*, 8(2):135–143, 2004.

[TS80] M Kamil Tulga and Thomas B Sheridan. Dynamic decisions and work load in multitask supervisory control. *Systems, Man and Cybernetics, IEEE Transactions on*, 10(5):217–232, 1980.

[TT05] Jianhua Tao and Tieniu Tan. Affective computing: A review. In *Affective computing and intelligent interaction*, pages 981–995. Springer, 2005.

[VC73] PH Venables and MI Christie. Mechanisms, instrumentation, recording techniques, and quantification of responses. *Electrodermal activity in psychological research*, pages 2–124, 1973.

[VLK99] Werner Van Leekwijck and Etienne E Kerre. Defuzzification: criteria and classification. *Fuzzy sets and systems*, 108(2):159–178, 1999.

[VTM87] Kim J Vicente, D Craig Thornton, and Neville Moray. Spectral analysis of sinus arrhythmia: A measure of mental effort. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 29(2):171–182, 1987.

[WP12] Wilhelm Wundt and Rudolf Trans Pintner. An introduction to psychology. 1912.

[WR10] Sherri C Widen and James A Russell. Descriptive and prescriptive definitions of emotion. *Emotion Review*, 2(4):377–378, 2010.

[YD08] Robert M Yerkes and John D Dodson. The relation of strength of stimulus to rapidity of habit-formation. *Journal of comparative neurology and psychology*, 18(5):459–482, 1908.

[YRS11] Michelle Yik, James A Russell, and James H Steiger. A 12-point circumplex structure of core affect. *Emotion*, 11(4):705, 2011.

[Zad65] Lotfi A Zadeh. Fuzzy sets. *Information and control*, 8(3):338–353, 1965.

[Zad96] Lotfi A Zadeh. Fuzzy logic= computing with words. *IEEE TRANSACTIONS ON FUZZY SYSTEMS*, 4(2):103, 1996.

[ZB06] Jing Zhai and Armando Barreto. Stress detection in computer users based on digital signal processing of noninvasive physiological variables. In *Engineering in Medicine and Biology Society, 2006. EMBS'06. 28th Annual International Conference of the IEEE*, pages 1355–1358. IEEE, 2006.

List of Figures

List of Tables

Listings

Acronyms

12-PAC 12-point affect circumplex. 24

ANN artificial neural network. 73

ANS autonomic nervous system. 7, 14, 15

API application programming interface. 70, 71, 81, 95, 100

BVP blood volume pulse. 47

CNS central nervous system. 5–7

ECG electrocardiogram. 12, 13, 16, 38, 40, 41, 43, 45–48, 56, 68

EDA Electrodermal activity. 7–10, 16, 20, 38, 40, 41, 45, 46, 50, 56, 66, 67, 69, 70, 72, 77, 79, 80, 84–86, 96–98, 104, 108, 109, 112, 116, 120, 121, 126, 127, 131, 147, 153, 155, 157–159, 173

EEG electroencephalography. 6, 20, 44, 45, 65, 72, 89

EMG electromyogram. 21, 38, 40, 41, 47, 48, 50, 65, 79

ERP event-related potentials. 6, 16

FCL Fuzzy Control Language. 101

FDA Food and Drug Administration. 66

FFT Fast Fourier Transformation. 99

FNIRS functional near infrafred spectroscopy. 43

FP Fisher projection. 42

GATT generic attribute profiles. 98

GEQ game experiencequestionnaire. 45

GLaDOS Genetic Lifeform and Disk Operating System. 50

GPS Global Positioning System. 61, 68, 70, 71, 155

GSR Galvanic Skin Response. 7, 38, 47, 48

HCI Human Computer Interaction. 1, 2, 22

HR heart rate. 38, 67, 70, 72, 79, 80, 84, 85, 96, 104, 108, 145, 155, 158

HRV Heart Rate Variability. 14–16, 42, 43, 56, 68, 72, 73, 79–81, 85, 89, 90, 92, 99, 104, 112, 125, 131, 145

LOD level of detail. 119–122, 126

LSTM Long Short Term Memory Neural Networks. 37, 38

NASA National Aeronautics and Space Administration. 67

NASA-TLX NASA Task Load Index. 16–18, 43, 123–125, 129, 130, 133, 135, 136, 138, 139, 141, 143, 145–147, 149, 150

NHL National Hockey League. 48

PET Positron emission tomography. 6

PNS peripheral nervous system. 5–7, 11

PSD power spectrum density. 77, 84, 85, 89, 90, 96, 98, 99

RESP respiratory. 47, 48

RNN recurrent neural networks. 38

SAL-DB Sensitive Artificial Listener Database. 37

SAM self-assessment questionnaire manikin. 20–22, 126, 127

SCL Skin Conductance Level. 9, 10, 20, 72

SCR Skin Conductance Response. 9–11

SDK software development kits. 95, 97

SFFS sequential floating forward search. 42

SFFS-FP sequential floating forward search with Fisher projection. 41, 42

SNS somatic nervous system. 7

SVM support vector machine. 30, 40–42, 73, 74

SVR Support-Vector Regression. 37

SWAT Subjective Workload Assessment Technique questionnaire. 18, 19

TEMP temperature. 47, 48

UAV unmanned aerial vehicles. 43, 44

Appendix

A. List of Publications

2015

Surfaces for Point Clouds using Non-Uniform Grids on the GPU. Schiffner D., Stockhausen C., Ritter M. (2015). Short papers proceedings WSCG2015, ISBN 978-80-86943-66-4, Plzen, Czech Republic.

2014

M. Pfeiffer, C. Stockhausen, D. Krömker (2014, October). "The InnocentButGuilty framework: a step towards GKT-enhanced applications". In: Proceedings of the 8th Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational (pp. 919-922). ACM.

2013

W. Müller, U. Spierling, C. Stockhausen (2013). "Production and Delivery of Interactive Narratives Based on Video Snippets". In: Interactive Storytelling. Springer International Publishing, 2013. 71-82.

C. Stockhausen, J. Smyzek, D. Krömker (2013). "Beats Down: Using Heart Rate for Game Interaction in Mobile Settings". In: Human-Computer Interaction INTERACT 2013. Springer Berlin Heidelberg, 2013. 523-530.

C.Stockhausen (2013). "StubSA: Studienbegleitende Self-Assessments in der Studieneingangsphase". In: E-Learning zwischen Vision und Alltag 2013, Medien in der Wissenschaft, Band 64

2012

K. Reitz, C. Stockhausen, D. Krömker (2012). ”Zone of impulse: physiological data enhanced gaming.” Proceedings of the 14th international conference on Human-computer interaction with mobile devices and services companion. ACM, 2012.

T. Föller, M.Koca, C.Stockhausen, H.Serve (2012). ”Entwicklung einer Smartphone Anwendung zur Unterstützung leitliniengerichteter Therapie von Tumorpatienten am Universitären Centrum für Tumorerkranungen Frankfurt”. GMDS 2012: 57. Jahrestagung der Deutschen Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie e. V. (GMDS)

C. Stockhausen, S. Voß, D.Weiß, D.Krömker (2012). ”Authoring and Training: Question Types for Interpretation and Comprehension of Images and Text”. In: Proceedings of the 5th International eLBa Science Conference in Rostock, Germany, June 21-22, 2012

M. Pfeiffer, C. Stockhausen, K. Reitz, D. Krömker (2012). ”Physiological Data in Future Living Environments”. In: ARCS Workshops (pp. 75-86).

2011

C. Stockhausen, D. Schiffner, D. Krömker (2011). ”Adaption of User Interfaces based on Physiological Data”. In: S. Schmid, M. Elepfandt, J. Adenauer, A. Lichtenstein (Eds.), 9. Berliner Werkstatt Mensch-Maschine-Systeme (pp. 78-79). Berlin: VDI.

2010

C. Stockhausen, D. Krömker (2010). ”Measuring Mental Load with a Polar Heart Rate Monitor”. In: R. Dörner, D. Krömker (Eds.), Proceedings of the first ITG/GI Workshop on Self Integrating Systems for Better Living Environments (pp. 45-49). Wiesbaden: Shaker Verlag.

B. Fuzzy Rule Sets

This appendix lists the different rules used for the single steps of the model in detail. The model is presented in chapter 6

B.1. Transformation of Arousal

Membership to one of the seven states for arousal (very high, high, mid high, medium, mid low, low and very low) is determined based on EDA and movement.

if (EDA is high) then (arousal is very high)
if (EDA is high and movement is not low) then (arousal is high)
if (EDA is mid high) then (arousal is high)
if (EDA is mid high and movement is not low) then (arousal is mid high)
if (EDA is medium) then (arousal is mid high)
if (EDA is medium and movement is not low) then (arousal is medium)
if (EDA is mid low) then (arousal is medium)
if (EDA is mid low and movement is not low) then (arousal is mid low)
if (EDA is low) then (arousal is mid low)
if (EDA is low and movement is not low) then (arousal is very low)

B.2. Transformation of Valence

Membership to one of the seven states for valence (very high, high, mid high, medium, mid low, low and very low) is determined based on heart rate, performance and movement.

IF HR is high THEN valence is high
IF HR is high AND Movement is high THEN valence is medium
IF HR is high AND (Performance is High OR mid High) AND movement is not high THEN valence is very high
IF HR is high AND (Performance is low OR mid low) AND movement is not high THEN valence is mid high
IF HR is high AND (Performance is High OR mid High) AND movement is high THEN valence is high
IF HR is mid high THEN valence is mid high
IF HR is mid high AND movement is high THEN valence is medium
IF HR is mid high AND performance is high THEN valence is high
IF HR is mid high AND (performance is low OR performance is mid low) THEN valence is medium
IF HR is medium THEN valence is medium
IF HR is medium AND movement is high THEN valence is mid low
IF HR is medium AND (performance is high OR performance is mid high) THEN valence is mid high
IF HR is mid low THEN valence is mid-low
IF HR is mid low AND movement is high THEN valence is medium
IF HR is mid low AND (performace is high OR performace is mid-high) THEN valence is mid-high
IF HR is low THEN valence is low
IF HR is low AND movement is not low THEN valence is very low
IF HR is low AND performance is low THEN valence is very low
IF HR is low AND (performance is high OR mid high) THEN valence is mid low

B.3. Transformation of Mental Effort

Membership to one of the four states for mental effort (low, medium, high, very high) is based on normalized power spectrum density results from HRV analysis.

IF PSD is low THEN mental effort is low
IF PSD is mid low THEN mental effort is medium
IF PSD is medium THEN mental effort is medium
IF PSD is mid high THEN mental effort is high
IF PSD is high THEN mental effort is very high

B.4. Transformation of Affective State

B.4.1. State 1 - Alarmed, Astonished

IF arousal is medium THEN state1 is very low
IF arousal is mid low THEN state1 is very low
IF arousal is low THEN state1 is very low
IF arousal is very low THEN state1 is very low
IF valence is very low THEN state1 is very low
IF valence is low THEN state1 is very low
IF valence is very high THEN state1 is very low
IF valence is high THEN state1 is very low
IF arousal is mid high AND valence is mid low THEN state1 is very low
IF arousal is mid high AND valence is mid high THEN state1 is very low
IF arousal is medium AND valence is medium THEN state1 is low
IF arousal is high AND valence is mid low THEN state1 is low
IF arousal is high AND valence is mid high THEN state1 is low
IF arousal is mid high AND valence is medium THEN state1 is medium
IF arousal is very high AND valence is mid low THEN state1 is medium
IF arousal is very high AND valence is mid high THEN state1 is medium
IF arousal is high AND valence is medium THEN state1 is high
IF arousal is very high AND valence is medium THEN state 1 is high

B.4.2. State 2 - Excited, Happy

IF arousal is very low THEN state2 is very low
IF arousal is low THEN state 2 is very low
IF arousal is mid low THEN state2 is very low
IF arousal is medium THEN state2 is very low
IF valence is very low THEN state2 is very low
IF valence is low THEN state2 is very low
IF valence is mid low THEN state2 is very low
IF valence is medium THEN state 2 is very low
IF arousal is medium AND valence is medium THEN state2 is low
IF arousal is mid high AND valence is medium THEN state2 is low
IF arousal is medium AND valence is mid high THEN state2 is low
IF arousal is very high AND valence is mid high THEN state2 is low
IF arousal is mid high AND valence is very high THEN state2 is low
IF arousal is mid high AND valence is mid high THEN state2 is medium
IF arousal is high AND valence is mid high THEN state2 is medium
IF arousal is mid high AND valence is high THEN state2 is medium
IF arousal is high AND valence is high THEN state2 is high
IF arousal is very high AND valence is high THEN state2 is high
IF arousal is high AND valence is very high THEN state2 is high
IF arousal is very high AND valence is very high THEN state2 is high

B.4.3. State 3 - Happy, Content

IF arousal is very low THEN state3 is very low
IF arousal is low THEN state3 is very low
IF arousal is mid low THEN state3 is very low
IF arousal is medium THEN state3 is very low
IF valence is very high THEN state3 is very low
IF valence is high THEN state3 is very low
IF valence is very low THEN state3 is very low
IF valence is low THEN state3 is very low
IF arousal is medium AND valence is medium THEN state3 is low
IF arousal is mid high AND valence is high THEN state3 is low
IF arousal is mid low AND valence is high THEN state3 is low
IF arousal is medium AND valence is mid high THEN state3 is medium
IF arousal is mid high AND valence is very high THEN state3 is medium
IF arousal is mid low AND valence is very high THEN state3 is medium
IF arousal is medium AND valence is high THEN state3 is high
IF arousal is medium AND valence is very high THEN state3 is high

B.4.4. State 4 - Relaxed, Calm

IF arousal is very high THEN state4 is very low
IF arousal is high THEN state4 is very low
IF arousal is mid high THEN state4 is very low
IF arousal is medium THEN state4 is very low
IF valence is medium THEN state4 is very low
IF valence is very low THEN state4 is very low
IF valence is low THEN state4 is very low
IF valence is mid low THEN state4 is very low
IF arousal is medium AND valence is medium THEN state4 is low
IF arousal is medium AND valence is mid high THEN state4 is low
IF arousal is mid low AND valence is medium THEN state4 is low
IF arousal is mid low AND valence is very high THEN state4 is low
IF arousal is very low AND valence is mid high THEN state4 is low
IF arousal is mid low AND valence is mid high THEN stat4 is medium
IF arousal is low AND valence is mid high THEN state4 is medium
IF arousal is mid low AND valence is high THEN state4 is medium
IF arousal is low AND valence is high THEN state4 is high
IF arousal is low AND valence is very high THEN state4 is high
IF arousal is very low AND valence is high THEN state4 is high
IF arousal is very low AND valence is very high THEN state4 is high

B.4.5. State 5 - Tired, Sleepy

IF arousal is medium THEN state5 is very low
IF arousal is mid high THEN state5 is very low
IF arousal is high THEN state 5 is very low
IF arousal is very high THEN state5 is very low
IF valence is very low THEN state5 is very low
IF valence is low THEN state5 is very low
IF valence is high THEN state5 is very low
IF valence is very high THEN state5 is very low
IF arousal is medium AND valence is medium THEN state5 is low
IF arousal is low AND valence is mid low THEN state5 is low
IF arousal is low AND valence is mid high THEN state5 is low
IF arousal is mid low AND valence is medium THEN state5 is medium
IF arousal is very low AND valence is mid low THEN state5 is medium
IF arousal is very low AND valence is mid high THEN state5 is medium
IF arousal is low AND valence is medium THEN state5 is high
IF arousal is very low AND valence is medium THEN state5 is high

B.4.6. State 6 - Bored, Depressed

IF arousal is medium THEN state6 is very low
IF arousal is mid high THEN state6 is very low
IF arousal is high THEN state6 is very low
IF arousal is very high THEN state6 is very low
IF valence is medium THEN state6 is very low
IF valence is mid high THEN state6 is very low
IF valence is high THEN state6 is very low
IF valence is very high THEN state6 is very low
IF arousal is medium AND valence is medium THEN state6 is low
IF arousal is medium AND valence is mid low THEN state6 is low
IF arousal is mid low AND valence is medium THEN state6 is low
IF arousal is very low AND valence is mid low THEN state6 is low
IF arousal is mid low AND valence is very low THEN state6 is low
IF arousal is mid low AND valence is mid low THEN state6 is medium
IF arousal is mid low AND valence is low THEN state6 is medium
IF arousal is low AND valence is mid low THEN state6 is medium
IF arousal is low AND valence is low THEN state6 is high
IF arousal is low AND valence is very low THEN state6 is high
IF arousal is very low AND valence is low THEN state6 is high
IF arousal is very low AND valence is very low THEN state6 is high

B.4.7. State 7 - Sad, Miserable

IF arousal is very high THEN state7 is very low
IF arousal is high THEN state7 is very low
IF arousal is very low THEN state7 is very low
IF arousal is low THEN state7 is very low
IF valence is medium THEN state7 is very low
IF valence is mid high THEN state7 is very low
IF valence is high THEN state7 is very low
IF valence is very high THEN state7 is very low
IF arousal is medium AND valence is medium THEN state7 is low
IF arousal is mid low AND valence is low THEN state7 is low
IF arousal is mid high AND valence is low THEN state7 is low
IF arousal is medium AND valence is mid low THEN state7 is medium
IF arousal is mid high AND valence is very low THEN state7 is medium
IF arousal is mid low AND valence is very low THEN state7 is medium
IF arousal is medium AND valence is low THEN state7 is high
IF arousal is medium AND valence is very low THEN state7 is high

B.4.8. State 8 - Frustrated, Angry

IF arousal is medium THEN state8 is very low
IF arousal is mid low THEN state8 is very low
IF arousal is low THEN state8 is very low
IF arousal is very low THEN state8 is very low
IF valence is medium THEN state8 is very low
IF valence is mid high THEN state8 is very low
IF valence is high THEN state8 is very low
IF valence is very high THEN state8 is very low
IF arousal is medium AND valence is medium THEN state8 is low
IF arousal is mid high AND valence is medium THEN state8 is low
IF arousal is medium and valence is mid low THEN state8 is low
IF arousal is very high AND valence is mid low THEN state8 is low
IF arousal is mid high AND valence is very low THEN state8 is low
IF arousal is mid high AND valence is mid low THEN state8 is medium
IF arousal is mid high AND valence is low THEN state8 is medium
IF arousal is high AND valence is mid low THEN state8 is medium
IF arousal is high AND valence is low THEN state8 is high
IF arousal is high AND valence is very low THEN state8 is high
IF arousal is very high AND valence is low THEN state8 is high
IF arousal is very high AND valence is very low THEN state8 is high

B.5. Transformation of Cognitive State

if (mentalLoad is very high) then (cognitiveState is very high)
if (mentalLoad is very high and performance is very high) then (cognitiveState is high)
if (mentalLoad is very high and performance is high) then (cognitiveState is high)
if (mentalLoad is high) then (cognitiveState is high)
if (mentalLoad is high and movement is high) then (cognitiveState is very high)
if (mentalLoad is high and performance is very low and movement is low) then (cognitiveState is very high)
if (mentalLoad is high and performance is very low and movement is not low) then (cognitiveState is very high)
if (mentalLoad is medium) then (cognitiveState is medium)
if (mentalLoad is medium and movement is high) then (cognitiveState is high)
if (mentalLoad is medium and performance is very low and movement is low) then (cognitiveState is low)
if (mentalLoad is medium and performance is very low and movement is high) then (cognitiveState is high)
if (mentalLoad is low) then (cognitiveState is low)
if (mentalLoad is low and movement is high) then (cognitiveState is medium)
if (mentalLoad is low and performance is high) then (cognitiveState is medium)
if (mentalLoad is low and performance is very high) then (cognitiveState is medium)