Department of Computer Science
Series of Publications A
Report A-2018-3

Implicit Interaction with Textual Information using Physiological Signals

Oswald Barral

To be presented, with the permission of the Faculty of Science of the University of Helsinki, for public examination in Auditorium XIII, University Main Building, on 1st of March 2018, at 12 o'clock noon.

University of Helsinki
Finland

Supervisors
Giulio Jacucci, University of Helsinki, Finland
Tuukka Ruotsalo, University of Helsinki, Finland

Pre-examiners
Elisabeth André, Augsburg University, Germany
Joemon M Jose, University of Glasgow, United Kingdom

Opponent
Anne-Marie Brouwer, Netherlands Organization for Applied Scientific Research TNO, Netherlands

Custos
Giulio Jacucci, University of Helsinki, Finland

Contact information

Department of Computer Science
P.O. Box 68 (Gustaf Hällströmin katu 2b)
FI-00014 University of Helsinki
Finland

Email address: info@cs.helsinki.fi
URL: http://cs.helsinki.fi/
Telephone: +358 2941 911, telefax: +358 9 876 4314

ISSN 1238-8645
ISBN 978-951-51-4070-8 (paperback)
ISBN 978-951-51-4071-5 (PDF)
Computing Reviews (1998) Classification: H.1.2, H.3.3, H.5.2
Helsinki 2018
Unigrafia

Implicit Interaction with Textual Information using Physiological Signals

Oswald Barral

Department of Computer Science
P.O. Box 68, FI-00014 University of Helsinki, Finland
oswald.barral@cs.helsinki.fi

PhD Thesis, Series of Publications A, Report A-2018-3
Helsinki, March 2018, 72+145 pages
ISSN 1238-8645
ISBN 978-951-51-4070-8 (paperback)
ISBN 978-951-51-4071-5 (PDF)

Abstract

Implicit interaction refers to human-computer interaction techniques that do not require active engagement from the users. Instead, the user is passively monitored while performing a computer task, and the data gathered is used to infer implicit measures as inputs to the system. Among the multiple applications for implicit interaction, collecting user feedback on information content is one that has increasingly been investigated. As the amount of available information increases, traditional methods that rely on the users' explicit input become less feasible. As measurement devices become less intrusive, physiological signals arise as a valid approach for generating implicit measures when users interact with information. These signals have mostly been investigated in response to audio-visual content, while it is still unclear how to use physiological signals for implicit interaction with textual information.

This dissertation contributes to the body of knowledge by studying physiological signals for implicit interaction with textual information. The research targets three main research areas: a) physiology for implicit relevance measures, b) physiology for implicit affect measures, and c) physiology for real-time implicit interaction. Together, these provide understanding not only on what type of implicit measures can be extracted from physiological signals from users interacting with textual information, but also on how these can be used in real time as part of fully integrated interactive information systems.

The first research area targets perceived relevance, as the most noteworthy underlying property regarding the user interaction with information items. Two experimental studies are presented that evaluate the potential for brain activity, electrodermal activity, and facial muscle activity as candidate measures to infer relevance from textual information. The second research area targets affective reactions of the users. The thesis presents two experimental studies that target brain activity, electrodermal activity, and cardiovascular activity to indicate users' affective responses to textual information.

The third research area focuses on demonstrating how these measures can be used in a closed interactive loop. The dissertation reports on two systems that use physiological signals to generate implicit measures that capture the user's responses to textual information. The systems demonstrate real-time generation of implicit physiological measures, as well as information recommendation on the basis of implicit physiological measures.

This thesis advances the understanding of how physiological signals can be implemented for implicit interaction in information systems. The work calls for researchers and practitioners to consider the use of physiological signals as implicit inputs for improved information delivery and personalization.

Computing Reviews (1998) Categories and Subject Descriptors:

H.1.2 [MODELS AND PRINCIPLES]: User/Machine Systems
H.3.3 [INFORMATION STORAGE AND RETRIEVAL]: Information Search and Retrieval—Relevance feedback
H.5.2 [INFORMATION INTERFACES AND PRESENTATION]: User Interfaces—Interaction styles, Input devices and strategies, Evaluation/methodology

General Terms:

Physiological Computing, Human-Computer Interaction, Information Retrieval

Additional Key Words and Phrases:

Physiological Text Annotation, Human-Centered Implicit Tagging, Implicit Relevance Feedback

Acknowledgements

This thesis is the result of close to four years of exciting research, which would not have been possible without the support and guidance of many people. This portion of the dissertation is dedicated to all of those who accompanied me during this intense journey.

This research received financial support from the Helsinki Doctoral Education Network in Information and Communications Technology (HICT) and from the European Commission through the Mindsee project. The Department of Computer Science at the University of Helsinki, and the Helsinki Institute for Information Technology (HIIT) both at Aalto University and University of Helsinki provided me with the infrastructure and ideal environment to conduct my research.

I share the credit of this work with my supervisor Professor Giulio Jacucci, who trusted me and provided the best conditions one could ask for to conduct quality research. Giulio gave me the freedom to guide my own research, and was always ready to provide advice and insightful comments on my work. Further, he gave me the chance to work with exceptional researchers not only in Finland, but from other European countries as well.

I extend my warmest thanks to my co-supervisor Tuukka Ruotsalo, and my mentors Manuel J.A. Eugster and Michiel Sovijärvi-Spapé. I have learned the most from them during this journey, and I could not have had such a great academic experience without their mentorship. From Tuukka I learned the scientific mindset, and to always keep a critical view on my work. Manuel passed a great deal of his knowledge of machine learning and statistics on to me over endless hours of data wrangling. Michiel taught me about the fascinating realm of human physiology and cognitive neuroscience.

I would like to thank Professor Elisabeth André, and Professor Joemon M. Jose for acting as pre-examiners of this thesis. Their rigorous and valuable feedback enabled the further strengthening of the contributions of my work. Furthermore, I am privileged to have Dr. Anne-Marie Brouwer as my opponent.

In addition to the above, I am indebted to all the co-authors of the articles that emanated from the research presented in this thesis: Benjamin Blankertz, Pedram Daee, Jonathan Freeman, Luciano Gamberini, Samuel Kaski, Patrik Pluchino, Niklas Ravaja, and Markus Wenzel. I am particularly grateful to Ilkka Kosunen, as we have worked hand in hand not only on several research ventures, but also on teaching and tutoring students. Further, I have had the privilege to be part of the Ubiquitous Interaction group, and would like to acknowledge its members as well as my colleagues and peers at the Department of Computer Science and at HIIT who have accompanied me throughout this journey: Imtiaj Ahmed, Salvatore Andolina, Kumaripaba Athukorala, Joanna Bergström-Lehtovirta, Diogo Cabral, Marco Filetti, Ville Harjunen, Chen He, Yi-Ta Hsieh, Antti Jylhä, Luana Micallef, Matti Nelimarkka, Antti Salovaara, Andrea Vianello, and Tung Vuong. Many of them have become close friends, especially Khalil Klouche and Barış Serim with whom I have shared office, lunch breaks, and countless after-work saunas. They are a fine example of how work can harmoniously blend with private life. Moreover, I thank the staff at the Department of Computer Science, notably the research coordinator Pirjo Moen for her kind support throughout these years. I extend my gratitude to research assistants Hassan Abbas and Nishadh Aluthge for their heroic effort in keeping up with me, and to the brave participants of my experiments for their willingness to provide immensely valuable research data. I also thank Brandon Startz for helping me to proofread this manuscript.

I am grateful to be surrounded by sincere friends and loving family, who have always been there when I needed them. In Finland, friends have been my pillar of strength, particularly Rahul and Sergio, with whom I have spent most of my time out of office. Back home, my loving parents, Clotilde and Xavier, have persistently encouraged me throughout the process of writing this thesis. My siblings, Astrid and Alexandre, have kept me close in their hearts and minds despite the distance, their continued support has been of inestimable value.

Final words of gratitude go to my late loved ones. I am deeply thankful to my friend Brusti for encouraging me to pursue my ambitions abroad. His positive attitude and fighting spirit have been, and will always remain, a true inspiration in times of adversity. Above all, I am indebted to my grandfather for teaching me patience, perseverance, humility, and commitment. This journey has challenged these values to the core, and that is why today, *l'Avi* would have been the proudest.

Helsinki, March 2018
Oswald Barral

List of Publications

The research presented in this thesis has led to the following six publications, which are referred to throughout the thesis as Publications I-VI. The publications are reproduced at the end of the thesis.

Publication I: Manuel J. A. Eugster, Tuukka Ruotsalo, Michiel M. Spapé, Ilkka Kosunen, Oswald Barral, Niklas Ravaja, Giulio Jacucci, and Samuel Kaski. Predicting term-relevance from brain signals. In *Proceedings of the 37th International Conference on Research & Development in Information Retrieval*, SIGIR '14, pages 425–434, 2014. ACM. [37]

Publication II: Oswald Barral, Manuel J. A. Eugster, Tuukka Ruotsalo, Michiel M. Spapé, Ilkka Kosunen, Niklas Ravaja, Samuel Kaski, and Giulio Jacucci. Exploring peripheral physiology as a predictor of perceived relevance in information retrieval. In *Proceedings of the 20th International Conference on Intelligent User Interfaces*, IUI '15, pages 389–399, 2015. ACM. [9]

Publication III: Oswald Barral, Ilkka Kosunen, Tuukka Ruotsalo, Michiel M. Spapé, Manuel J. A. Eugster, Niklas Ravaja, Samuel Kaski, and Giulio Jacucci. Extracting relevance and affect information from physiological text annotation. *User Modeling and User-Adapted Interaction*, 26(5):493–520, 2016. [12]

Publication IV: Oswald Barral, Ilkka Kosunen, and Giulio Jacucci. No need to laugh out loud: Predicting humor appraisal of comic strips based on physiological signals in a realistic environment. *ACM Transactions on Computer-Human Interaction.* 24(6):40, 2017. [11]

Publication V: Manuel J. A. Eugster, Tuukka Ruotsalo, Michiel M. Spapé, Oswald Barral, Niklas Ravaja, Giulio Jacucci, and Samuel Kaski. Natural brain-information interfaces: Recommending information by relevance inferred from human brain signals. *Scientific Reports*, 6(1):38580, 2016. [38]

Publication VI: Giulio Jacucci, Oswald Barral, Pedram Daee, Markus Wenzel, Baris Serim, Tuukka Ruotsalo, Patrik Pluchino, Jonathan Freeman, Luciano Gamberini, Samuel Kaski, and Benjamin Blankertz. Integrating neurophysiological relevance feedback in intent modeling for information retrieval. *Submitted to Journal* [60]

Contents

Chapter 1

Introduction

Implicit interaction refers to human-computer interaction (HCI) techniques that support user interaction without active engagement of the users. Implicit interaction often relies on monitoring the user behavior and interaction history in order to infer measures that can be used as inputs to information systems. One of the most appealing characteristics of implicit interaction techniques is that they enable input to the systems at no additional cost from the user. The increasing availability and amount of information content poses challenges to information access, and it becomes harder to collect user feedback describing how users respond to information items. The traditional way of collecting such feedback has been to rely on the users' explicit input, examples include profiling questionnaires in user modeling [68], relevance feedback in information retrieval [106], and manual annotation for content tagging [4]. Inferring implicit measures representing the user feedback (also known as implicit feedback) has increasingly been investigated to replace or complement explicit user feedback, as users have proved not to be willing to interrupt their task in order to provide feedback [66]. For instance, YouTube recommendations are not only based on the users' likes or subscriptions to channels (*explicit*), but also take into account the time users spend watching videos (*implicit*) [35].

Implicit measures present the advantage that they can be collected at a high throughput, as they do not rely on the explicit engagement of the user. However, they are less accurate than explicit measures as they are inferred measures, and thus may fail to capture the true response from the user. Following the previous example, a user might be playing a video, but not paying attention to it, which would not be captured using solely "playing time" as an implicit measure of user interest.

Physiological signals have been studied in psychology as they have been shown to reflect our moods and behaviors [28]. As devices that allow physiological recordings are becoming more affordable and less intrusive, the field of physiological computing, which concerns the use of these signals for real-time interaction with computing systems, is growing fast [40, 41, 61]. One characteristic that makes physiological signals especially interesting is that they can potentially reveal the user responses in real time, and at no additional interaction cost. These attributes make them especially well-suited for the problem of generating implicit measures to infer user feedback on information items. The use of physiological signals for implicit interaction with information is still for the most part restricted to images and video content. However, it is not yet understood how physiological signals can be used to infer implicit measures of the users' responses to textual information. Further, the generation and use of such measures in real time has yet to be investigated.

1.1 Objectives and Scope

The goal of this thesis is to investigate physiological signals for implicit interaction with textual information. This involves not only investigating these signals to generate measures to gather user feedback, but also evaluating their use for implicit interaction in real time. In order to provide a comprehensive approach to the research problem, the objectives are broken down into three main research areas, which are introduced below.

Investigate the potential of physiological signals to indicate measures of the user's perceived relevance of textual information: One important underlying property often used in information systems regarding the users' interaction with information items is the perception of relevance. That is, whether the users find a piece of information relevant for their information needs. In information retrieval, implicit behavioral and physiological sources have been explored to infer relevance of information, namely dwell time and click-through activity [67], eye movements [24, 105], pupil size [89], and facial expressions [7]. While some work has begun to consider physiological signals as implicit measures for relevance (e.g., brain signals for image retrieval [64]), the use of these signals to infer measures targeting the relevance of textual information is still largely unexplored.

Investigate the potential of physiological signals to indicate measures of the user's affective responses to textual information: Affect refers to the feeling or expression of an emotion, and physiological signals have increasingly been used to indicate users' affect in HCI setups [33]. While research has tackled physiological signals as implicit channels to infer affective clues of the users during information consumption setups, these are mostly restricted to images and video content. Further, textual information consumed in everyday situations is often of a mixed nature. For instance, a news article might be accompanied by images, and it is unclear how physiology can be used to indicate affective cues on this type of content. Specifically concerning the field of information systems, emotional and affective cues have been studied in what is known as affective recommender systems [120]. These systems use inferred emotions through affective features as contextual information for generating the recommendations [50]. Typically, affective information is extracted from the information content, from contextual and behavioral clues, or other signals such as facial expressions, but physiological signals have yet to be explored in this domain.

Investigate the use of physiological signals for real-time implicit interaction with textual information: Physiological computing systems are defined by their ability to provide real-time adaptation based on the recorded physiological signals [40]. However, when using physiological signals to implicitly collect user feedback on information, the body of research mostly concentrates on offline analysis approaches. That is, data is collected under certain experimental conditions, and physiological signals are later analyzed in order to infer measures describing the users' interaction with information. It is a main objective of this thesis to target real-time adaptation, validating the offline extraction of implicit measures from physiological signals for implicit interaction with information systems.

1.2 Research Contributions

The research presented in this thesis contributes to the body of knowledge by advancing the understanding of how physiological signals can be used for implicit interaction with textual information in interactive information systems. The greater contribution of this thesis is that it provides a broad understanding of the topic, as enabled by bringing together six publications that originated from the research process. Overall, this thesis contributes to the body of knowledge with the following three main contributions, one in each targeted research area:

C1: Demonstrating the use of physiological signals to infer the user's perceived relevance of textual information (Publications I and II)

C2: Demonstrating the use of physiological signals to infer the user's affective responses to textual information (Publications III and IV)

C3: Demonstrating the use of physiological signals for real-time implicit interaction with textual information (Publications V and VI)

The contributions of each of the six publications separately, as well as the individual role of each author involved, are detailed below.

Publication I – Predicting term-relevance from brain signals
In this publication we demonstrate the use of brain signals alone to predict perceived relevance of terms. We present an experimental study and machine learning models to demonstrate that brain signals can be used as a unique source to infer relevance judgments from individual terms significantly better than a random baseline. In addition, we identify where and at which point in time relevance signals are best captured in the measured brain signals.

Individual contributions: The data recording experiment was designed by Ilkka Kosunen, with contributions from the rest of the authors. I implemented the experimental setup and ran the user studies, as well as wrote the corresponding sections in the paper. Manuel J.A. Eugster, Tuukka Ruotsalo and Michiel M. Spapé share the contribution regarding data analysis and paper writing. All of the other authors, in minor degree, contributed to the writing, and all the authors of the paper participated in the revisions. Samuel Kaski and Giulio Jacucci supervised the process.

Publication II – Exploring peripheral physiology as a predictor of perceived relevance in information retrieval
In this publication we explore facial muscles and skin conductance as physiological indicators of relevance of text snippets. We provide statistical analyses to identify where relevance decisions are most reflected in the physiological signals. Also, we present machine learning models to demonstrate that we are able to predict relevance judgments from the physiological signals significantly better than a random baseline.

Individual contributions: The data recording experiment was designed by Ilkka Kosunen, with contributions from the rest of the authors. I implemented the experimental setup and ran the user studies. I ran the physiological analysis with supervision from Michiel M. Spapé. Manuel J.A. Eugster designed the machine learning setup, and I implemented the feature selection. I led the paper writing, with input from Tuukka Ruotsalo, Manuel J.A. Eugster and Michiel M. Spapé. Giulio Jacucci supervised the process, and all of the authors participated in the revisions.

Publication III – Extracting relevance and affect information from physiological text annotation

In this publication we present the physiological text annotation framework, which explores peripheral physiology for generating metadata on information items related to relevance and affect. The framework is demonstrated by including the experimental setup and partial results of Publication II, as well as an additional experiment investigating skin conductance as a measure to indicate affective responses to news articles.

Individual contributions: The experiment design and implementation of the news articles experiment was done by me and Ilkka Kosunen. I ran the experiment, and generated the features from the data, while Ilkka Kosunen ran the data analysis. I led the paper writing, with significant input from Ilkka Kosunen, as well as Tuukka Ruotsalo, and Michiel M. Spapé, supervised by Giulio Jacucci. All the authors contributed to paper revisions.

Publication IV – No need to laugh out loud: Predicting humor appraisal of comic strips based on physiological signals in a realistic environment

In this publication we investigate brain activity, cardiovascular measures, and skin conductance to detect humor appraisal. We present an experimental setup and thorough analysis to evaluate physiological measures for automatic detection of humor appraisal. Predictive models are presented to demonstrate the predictive power of the physiological signals, as well as compared to state-of-the-art facial detection analysis from video recordings. In addition, we provide detailed analysis on how humor appraisal is captured in the physiological recordings.

Individual contributions: The experimental setup and data analysis procedures were jointly designed by me and Ilkka Kosunen. I carried out the data collection and analysis. The contribution to the writing is shared between me and Ilkka Kosunen, under the supervision of Giulio Jacucci.

Publication V – Natural brain-information interfaces: Recommending information by relevance inferred from human brain signals

In this publication we demonstrate the use of brain signals alone to recommend new information to users based on the relevance of terms inferred from brain signals only. We present an experiment in which brain signals of participants are recorded while they read a series of articles in order to infer term relevance measures. We evaluate new article recommendations to the users based on the inferred relevance measures, demonstrating the first-of-its-kind system that is able to recommend new valuable information to users based on their brain signals only.

Individual contributions: Manuel J.A. Eugster, Tuukka Ruotsalo and Michiel M. Spapé share the major contribution regarding system design, data analysis, and paper writing. I ran the experimental studies. All of the authors of the paper participated in the revisions. Niklas Ravaja, Giulio Jacucci, and Samuel Kaski supervised the process.

Publication VI – Integrating neurophysiological relevance feedback in intent modeling for information retrieval

In this publication, we demonstrate real-time generation of relevance measures from neurophysiological signals. We present a fully integrated information retrieval system that uses measures of relevance as directly inferred from brain activity and eye movements to estimate the users' search intentions. We present an evaluation experiment that allows to demonstrate real-time generation of relevance measures for implicit interaction paradigms.

Individual contributions: The system implementation and experiment design was a joint effort from the various authors of the paper, led by Giulio Jacucci. Markus Wenzel and Benjamin Blankertz implemented the neurophysiology-based relevance predictor. Pedram Daee, under the supervision of Samuel Kaski, implemented the user intent model and information retrieval system. I, with the close assistance of Baris Serim, led the implementation of the interface and integration of the several components, which was materialized by research assistants. The data collection was carried out by Patrik Plucchino, supervised by Luciano Gamberini, and research assistants supervised by me and Baris Serim. I carried out the data analysis, with input from Baris Serim and Pedram Daee. Giulio Jacucci and me led the paper writing process. Benjamin Blankertz, Pedram Daee, Tuukka Ruotsalo, and Baris Serim, provided input for major sections of the paper. All the authors contributed to paper revisions.

1.3 Structure of the Thesis

The thesis is organized as follows. Chapter 2 provides preliminaries on the physiological signals used in this thesis, followed by an overview of the most closely related research literature, allowing the identification of open challenges regarding physiology for implicit interaction with textual information.

Chapter 3 presents the research questions addressed in this thesis to tackle the identified open challenges. The chapter also provides insight into the research methods implemented to answer the research questions. Ethical concerns tied to the proposed research are discussed as well.

Chapter 4 presents the research carried out addressing physiological signals for inferring measures of relevance of textual information. The chapter overviews two experimental studies presented in Publications I and II, highlighting their joint contributions through their main results and findings.

Chapter 5 reports on the research carried out in using physiological signals to infer affective responses of the users to textual information items. The chapter presents two experimental studies and their main results and findings, as reported in Publications III and IV, as well as highlights their joint contributions.

Chapter 6 presents the research carried out for using implicit measures from physiological signals for real-time interaction with textual information. The chapter presents two systems that demonstrate real-time implicit interaction with textual information, as reported in Publications V and VI.

The main contributions of the thesis are summarized in Chapter 7, and their implications to the field are discussed. In addition, limitations of the proposed research are identified, pointing out directions for future work. The thesis is then wrapped up with concluding remarks.

Chapter 2

Background

This chapter provides necessary background information for the reader. First, the chapter provides preliminary definitions in psychophysiology, introducing what the physiological signals used in this thesis are, and how they are measured. Then, the chapter overviews the most closely related work that allows the identification of the open issues yet to be addressed in the body of work, which are the motivating force behind the research presented in this thesis.

2.1 Preliminaries of Psychophysiology

Psychophysiology refers to the science that studies how bodily reactions are related to our moods, behavior, sentiments, elations, and frustrations [28]. This thesis mainly focuses on physiological signals for their use in HCI setups, targeting mainly the central nervous system (CNS) and autonomic nervous system (ANS). Within the CNS, this thesis makes use of electroencephalography (EEG) as a method to measure brain-activity. Within the ANS, this thesis utilizes electrodermal activity (EDA) to capture sweat activity, electroencephalography (ECG) to capture cardiovascular activity, and electromyography (EMG) to capture muscle activity (see Figure 2.1). Provided below are further details on each of these signals, together with a brief overview on how they have been used in HCI setups. For a thorough overview of physiological measures in HCI, the reader is referred to our publication *The Psychophysiology Premier* [33], which has not been included in this thesis.

Figure 2.1: Physiological sensors used in this thesis. Electroencephalography (EEG, top-left), electrodermal activity (EDA, top-right), facial electromyography (fEMG, bottom-left), and electrocardiography (ECG, bottom-right).

2.1.1 Electroencephalography

Electroencephalography (EEG) measures the electrical activity on the scalp, generated by the activation of the neurons. Common EEG measurement devices involve placing several electrodes on the user's scalp (ranging from 1 to 256, but typically 32 or 64), which together with conductive gel, allow to capture electrical brain activity. High-end EEG amplifiers typically measure brain activity at 1000 - 2000 Hz. EEG is of special interest for HCI, as it allows for brain activity recording in a non-invasive manner (e.g., it does not involve surgical procedures), and it is light-weight (as opposed to other brain-measurement devices that involve room-sized sensors). However, while within the range of brain-measurement devices EEG can be considered non-intrusive, high quality EEG recordings are still arguably cumbersome, involving $40 - 60$ minutes to prepare a standard 32-channel

EEG recording setup. There are existing commercial devices that allow for easy and light-weight measuring of EEG, and studies have proven that these are valid approaches for certain HCI applications [73, 133]. Also, there is research targeting miniaturization of EEG sensors, in order to allow continuous measurement of EEG activity in a non-intrusive manner [88].

Brain-computer interfaces (BCIs) refer to systems that involve brain measures to interact with machines. BCIs can be divided into active BCIs, which involve explicit engagement of the users, and passive BCIs, which refer to passively monitoring the users' brain activity while they carry out HCI tasks in order to infer properties of their state [134]. Active BCIs have traditionally been studied as input control for motor-impaired patients to interact with machines, example applications include text spellers [74], and wheel-chair control [78]. Passive BCIs on the contrary target the general population, by adding brain activity measurement as an additional communication channel between human and machines, example applications include workload monitoring [20, 48], assisted car-braking [58], and image search [49].

2.1.2 Electrodermal Activity

Electrodermal activity (EDA), also known as galvanic skin response (GSR), measures the changes in the electrical properties of the skin, due to the varying level of sweat-induced moisture. A common method to measure EDA is to place two surface electrodes on two different points of the user's skin (e.g., middle phalanges of two different fingers), and a small, unnoticeable current is passed through them in order to measure the electrical resistance of the skin [17]. EDA has commonly been used to measure the activation of the sympathetic nervous system; therefore, it has proved to be a good indicator of the level of psychological, physiological and emotional arousal [6]. The EDA signal can be roughly separated into two components: *phasic*, and *tonic* activity. The former refers to electrodermal responses to stimuli and is characterized with spikes in the signal (i.e., skin conductance response; SCR), whereas the latter represents the base skin conductance level which is more related to ambient factors such as temperature, moisture, etc. [17].

EDA can be measured in a non-intrusive manner, and the *phasic* activity is not only a well-established measure of arousal, but has also been associated to stimulus novelty, intensity, and emotional content [33, 102]. The signal is very well suited for HCI scenarios as it can easily be measured using non-invasive wearable technology. For instance, studies have shown the use of wearable EDA sensors to detect stress [111], and wrist-bands [80] or rings[1] have been envisaged to measure EDA in wearable setups. Further, EDA is a reasonably simple signal, which does not require large computational resources to be processed. Also, its temporal nature is low, and therefore measuring EDA at 60Hz is enough for most applications.

2.1.3 Electromyography

Electromyography (EMG), measures myoelectric potentials generated by muscle activation using surface electrodes [27]. As the frequencies of interest in EMG signals can be up to 500Hz, it is preferred to record the signal at a sampling rate of at least 1000Hz [27, 33]. Electrodes can be placed on any muscle group in order to measure their activity. EMG sensors can also be integrated in wearable technology, example commercial applications include EMG sensors on sportswear to quantify the effect of muscle training[2].

In psychophysiology, muscles that are especially interesting are facial muscles, which are captured using facial electromyography (fEMG). Some of these facial muscles are of special interest as they are mostly involuntary, and have shown to reflect affective dimensions of the user state. For instance, *orbicularis occuli* (muscle below the eye), together with *zygomaticus major* (cheek muscle), are believed to indicate a “true smile”, that can’t be faked [102]. On the other hand, *corrugator supercilii* (brow muscle) has been linked to negative emotional valence as it reflects frowning activity [102]. Thereafter, fEMG is commonly studied for indicating emotional valence [26]. Further, *corrugator supercilii* has been linked to other cognitive states and processes such as mental workload [122], fatigue [124], and compensatory mental effort [128].

In HCI, EMG has mostly been used as direct control input, applications ranging from the use of forearm muscles for gesture control [107, 136] to the control of prosthetics [14, 29]. While some effort has been put into studying EMG and especially fEMG as additional implicit interaction inputs to other HCI setups (e.g., video games [75]), this area is still largely unexplored.

2.1.4 Electrocardiography

Electrocardiography (ECG) captures the electrical activity of the heart, using chest electrodes [28]. In psychophysiology, cardiovascular measures are typically analyzed in the time-scale of minutes, and therefore recording

[1] `www.moodmetric.com`
[2] `www.liveathos.com`

does not require high sampling rates. Most commonly, inter-beat interval series are extracted from the raw ECG signal, and variability metrics (such as heart-rate variability) are used as indexes of cognitive states and processes [33].

ECG is only one of the possible techniques to measure cardiovascular activity, the most commonly used methods also include blood pressure (BP) [33], and photoplethysmography (PPG) [3]. Cardiovascular measures in psychophysiology have been mostly investigated as indicators of mental workload [1], mental stress and sympathetic and parasympathetic nervous system activity [21, 22, 33, 123]. Wearable technology to measure cardiovascular signals using ECG include chest straps [117], or wrist-worn devices [56]; while most of the modern smart-watches include PPG sensors for heart rate monitoring. Research has also considered remote measuring of cardiovascular measures using webcams [79], or cameras on mobile phones [55].

In HCI, cardiovascular measures are mostly used for sport [94] and health tracking [34, 90], but have also been studied for workload inference [132], and affective interfaces [45].

2.2 Physiology for Implicit Relevance Feedback

Relevance feedback consists of gathering relevance assessments from users when they are examining specific information items. The relevance assessments can then be utilized in feedback loops to specify the user's information need in subsequent iterations [57], to direct a search using visual interfaces [104, 116], or to gather relevance assessments from users for evaluation purposes [65]. One way to obtain relevance assessments is by direct input from the users, also known as explicit relevance feedback. Despite the robustness of this method, it is operationalized at the expense of the users' cognitive resources and manual effort [65]. Eventually, as the task complexity increases the cognitive resources required from the users, the technique becomes insufficient due to the cognitive burden that it causes [66].

Another way to obtain the relevance assessments is through implicit interaction, passively observing the users as they interact with the system [67]. This method is known as implicit relevance feedback, and it has been implemented mostly either through the use of surrogate measures based on interaction with information items (e.g., dwell time, click-through activity) or using other interaction data. Implicit feedback is not as robust as explicit feedback techniques [131] but, on the other hand, these measures have the advantage of coming at high throughput and with no additional cost from the user.

Novel inputs that provide user data which is not available through conventional channels are being increasingly investigated for implicit relevance feedback. For example, eye movements and pupil dilation have been investigated as implicit relevance measures, by inferring users' interest [25, 52, 89]. Facial expressions through video-recordings have also been explored for relevance inference [7], even in combination with physiological measures [8, 84]. As a matter of fact, in recent years, neuro-physiological measures have been drawing increasing attention in information retrieval [49, 53, 54, 86]. For instance, Kauppi et al. (2015) studied magnetoencephalographic signals alone and together with gaze signals in order to provide relevance feedback in an image retrieval task, using a static image database [64]. Similarly, Golenia et al. (2015) used EEG signals in combination with eye tracking to help users disambiguate image search results, in real time [49]. Very recently, Wenzel et. al (2017) used the same approach to predict relevance from keywords in a relevance assessment task in order to infer in real-time a given category of interest [130].

2.3 Physiology for Implicit Affective Annotation

Annotation refers to adding descriptive metadata to digital content and has traditionally been a critical backbone of many digital media services. It allows content management and analysis by enabling additional detailed information about the content. These annotations often rely on human input, be it self-reported, expert users, or crowds. Examples include bookmarking, rating, and tagging. To a certain level, these are digital adaptations of traditional pen-based annotations. Similarly as with explicit relevance feedback, it is operationalized at the expense of the users' cognitive resources and manual effort. Thus, with the increase of information available and user-generated content, alternative methods which do not rely on active engagement of the users are becoming the only viable and scalable solution. For instance, human-centered implicit tagging is one type of annotation which considers the users' natural responses to the information content [112, 127]. Implicit tagging concerns the users' reactions directly, as typically measured through facial features or physiological recordings. These implicit measures can be used for instance to improve content-based annotation [63, 70, 113].

Nowadays, especially within social media, users annotate content with affective cues, which represent their feeling or expression of an emotion [118].

However, self-reported affective annotations have shown to be problematic as they do not always reflect the real affective reaction of the users, but rather the emotions that users want to socially communicate [96]. The role of emotions in information sciences has been studied extensively [81, 82]; and physiological signals have been used as additional channels to implicitly measure human emotions and map affective states [31, 59, 69, 91]. Thus, using physiological signals to generate implicit affective annotations (or tags), rise as a potential alternative or complement to self-reported annotations. For instance, Koelstra et al. (2013) used a combination of facial expression recognition and physiological signals (namely EEG) to generate implicit affective tags on videos, in the arousal-valence space [70]. Similarly, Soleymani and Pantic (2013) used a reduced set of EEG electrodes to not only generate affective tags on videos, in the arousal-valence space, but also to detect relevant and irrelevant tags that had been pre-assigned to images [113].

Further, these affective annotations have proven to be useful for systems that use affective metadata to better support the recommendation of results to their users [119, 121]. These are known as affective recommender systems, which use affective features to infer the user's emotional responses to information items, in turn using them as contextual information for generating the recommendations [50]. For example, Tkalcic et al. (2010) studied image recommender systems using affective and generic metadata for improving image recommendation results, showing that including affective metadata compared to using generic metadata alone leads to better performance of the recommender system [119].

2.4 Open Challenges

This section has first introduced the physiological signals used in this thesis. This provided the reader with preliminary information regarding how these measures are recorded and what they reflect, but also in which HCI setups these signals are mostly being used. Further, the most closely related work to the proposed research has been overviewed. In light of the above, I identified the following main open challenges, which have motivated the research carried out in this thesis around the three main research areas presented in Section 1.1.

Lack of understanding of physiological signals in response to textual information. Most of the overviewed research targeting physiological signals for relevance inference or for affect inference is studied in response to image or video content. Using this type of content allows the measurement of users for extended periods of time, which in itself is known to cause emotional responses and more substantial physiological responses [84, 85]. However, while information items that consist of textual content (either in their totality, or partially) represent a great portion of the totality of information items consumed, these have only marginally been studied in the literature (e.g., [19]).

Lack of understanding of physiological signals to indicate relevance of information. The overviewed recent work using physiological signals to infer relevance measures do not target textual content, and is mostly studied in combination with other implicit measures such as video recordings. Therefore, at this stage, it is difficult to assess the capabilities of physiological signals to infer relevance of textual information. Different types of textual information might present different physiological responses, which may be better captured using one or another signal. It is still unclear which signals are the most useful and how they should be used (e.g., what are the best time windows for physiology-based relevance prediction) depending on the type of stimulus (words, text snippets, documents, etc.).

Lack of understanding of physiological signals to indicate affect of non-controlled textual information. Most of the research targeting physiological responses to information items are carried out under constrained laboratory setups, using controlled datasets. These have the advantage that the content presented to the users is known, which facilitates the experimental design and analysis. However, when addressing real-world information systems that for instance consider user-generated content, the information items are highly heterogeneous, which makes it harder to know beforehand to which affective category the item belongs (i.e., "ground truth"). Further, textual information might be accompanied with other types of content such as images, and it is mostly unclear how to investigate physiological signals for affect inference under these types of settings.

Lack of understanding of how to use physiological signals for real–time interaction with information. The field that deals with physiological signals for real-time adaptation is physiological computing [40, 41]. In these systems, physiological responses are measured in order to infer properties of the users, which are then fed back in real time to the system, leading to system adaptations which again influence users' physiological responses. This is referred to as the bio-cybernetic loop [39, 99]. Applications of physiological computing systems are broad, examples including improved task allocation for aviation operators [99, 100], video-games [75], or aided meditation [71]. BCIs have for long been investigated as control input channels especially targeting motor-impaired users [87]. However, research targeting physiological signals for interacting with information mainly focuses on inferring cognitive properties of the users. In other words, there is a lack of research investigating physiological signals to provide real-time interaction with information, closing the bio-cybernetic loop.

Chapter 3

Research Questions and Methods

The purpose of this thesis is to explore implicit interaction paradigms using physiological responses to textual information items. Chapter 1 has introduced the three main areas in which the research is centered. Following, Chapter 2 has overviewed the most closely related work, which allowed the identification of the major open challenges present in the literature. In order to address these open issues, the thesis has been articulated around three research questions (RQs). The RQs do not exhaustively address all aspects of the main research problem, as their function is to demarcate the research so that it is possible to conduct within the scope of a PhD thesis.

This section first introduces the RQs which are the backbone of the presented research, and how the publications compiled in this thesis provide answers to them. Following, the research strategies and methods used to address the RQs are presented and discussed. Finally, ethical concerns of the research and how they are addressed in this thesis are discussed.

3.1 Research Questions

RQ1: *Can physiological signals be used for implicit relevance inference of textual information?* When interacting with an information item, the most important underlying property is how relevant the item is perceived to be by the user. The need for gathering measures of how relevant users perceive the information items to be grows as the amount of information managed by information systems increase. Physiological signals have been investigated as potential sources for indicating relevance of information items. However, their study has mostly been limited to images and video content, and it remains unclear how physiological signals can be used to generate implicit measures of relevance of textual information.

Publications I and II report on two experimental studies that target several physiological signals to infer relevance of text items. Physiological signals are recorded when users perform relevance assessment tasks. Classification models are presented that generate implicit relevance measures of information items from physiological signals. In addition, the publications investigate how and at which point in time relevance judgments are reflected on the physiological signals. Together, the publications present evidence for implicit relevance inference through physiological measures.

RQ2: *Can physiological signals be used for implicit affect inference of non-controlled textual information?* Once an item is deemed relevant, it is consumed, possibly eliciting affective responses. Detecting these affective responses provides important clues to the users' preferences. Physiological signals are well suited for implicit affect detection as they have been related to the users' moods and behaviors, and can rarely be faked. However, while physiological signals have been mostly studied to indicate affective states in response to different types of controlled audio-visual content, it is still not clear how they can be used to infer affective states in response to non-controlled and heterogeneous textual information.

Publications III and IV report on two experimental studies that investigate physiological signals of users consuming non-controlled information items, with the goal of inferring measures of their affective responses. Physiological signals are recorded while users browse information items that are fetched in real time from the web. Together, the publications present evidence for the use of physiological signals to generate implicit measures of affect to textual information commonly found in everyday situations.

RQ3: *Can physiological signals reliably be used for real-time implicit interaction with textual information?* Building computational models that use physiological measures to infer properties of the information items being consumed is essential towards the validation of these signals for such use. However, applying these models to working systems that provide real-time adaptation adds an additional complexity layer that needs to be investigated. First, novel interaction paradigms that enable the elicitation of physiological responses without disrupting the users' natural behavior need to be investigated. Further, the use of real systems in real situations involve an intrinsic loss of control over confounding factors as compared to experimental studies, compromising the generalizability of the results obtained under controlled experimental conditions.

Publications V and VI report on two experimental systems that demonstrate the use of physiological signals for real-time implicit interaction with textual information. Publication V focuses on demonstrating the benefits of implicit physiological measures for improved information filtering. Further, Publication VI focuses on demonstrating the generation of implicit measures from physiological signals in real time. Together, the publications present a fundamental understanding for real-time implicit interaction with textual information using physiological signals.

3.2 Research Methods

The thesis uses quantitative research methods to address the aforementioned research questions. Experimental research methods are used to collect research evidence. The data is then analyzed using statistical methods, with a primary focus on machine learning approaches. Detailed below are the research methods used in this thesis regarding data collection and analysis.

As identified in Chapter 2, there is a lack of research literature targeting physiological signals in response to *textual information items*. Accordingly, this research focuses on collecting physiological responses to information items in the form of textual content. To elaborate, when addressing RQ1 and RQ3, physiological responses are studied to infer relevance of information items that solely consist of text. On the other hand, when studying affective reactions to information items in addressing RQ2, the approach is to lessen the degree of control in the experimental conditions, by using information items fetched in real time from the web. In these cases information items are often of mixed nature, for instance, composed of textual information accompanied by images.

McGrath (1995) defines three criteria that should be maximized when collecting research evidence: the generalizability of the collected evidence across the population, the precision and control over the measured behaviors and confounding factors, and the realism of the context in which the evidence is collected [83]. As explained by the author, by construct, it is impossible to maximize the three criteria at once, and thus it is left to researcher discretion to evaluate what is the most pertinent balance between the three criteria for the specific research problem being investigated.

Throughout this thesis, the chosen methods for collecting research evidence generally lean towards *maximizing the precision*. That being said, the physiological sensors used in the research are selected so that they can potentially be used for real-world interactive systems, leaving out possibly more precise measurements that would not be feasible in such interactive scenarios (e.g., room-sized sensors such as fMRI). While physiological sensors are becoming less obtrusive, in many cases high-quality recordings still rely on - often cumbersome - sensor setups, which restrict their use outside of the laboratory. This narrows the possibilities of carrying out field studies (that would maximize realism) or large sampling surveys (which would maximize generalizability). The research presented in this thesis mostly uses high-end physiological amplifiers to record physiological signals. In addition, lighter physiological recording setups that resemble future wearable setups are also investigated. For instance, when addressing RQ2, physiological signals are recorded using light-weight recording devices (e.g., avoiding full EEG measurements, which are more cumbersome), which increased realism, while using high-quality physiological amplifiers, which still maximized precision.

Laboratory experiments are used for collecting research evidence when targeting RQ1, which is a research strategy that maximizes precision. Laboratory experiments allow to control for as many confounding factors as possible, at the expense of the realism of the setup and task. As the relevance of textual information items is not yet very well understood, it is important in the research design to maximize the control over as many confounding factors as possible, in order to better investigate the physiological correlates of perceived relevance of information items [23]. However, some confounding factors, while being reduced using controlled laboratory experiments, can't be completely avoided. For instance, the user might remember that she forgot a dentist appointment, eliciting a physiological response that is likely to pollute the recorded physiological signals. The research design has to be able to mitigate such factors, for instance using repetitions and randomization [44].

Moving away from restricted laboratory experiments, towards more realistic information interaction setups, *experimental simulation* is used to address RQ2. The focus here is to investigate whether physiological signals could be used to infer affective properties of information items in more naturalistic scenarios. This research strategy involves "concocting a situation or behavior setting or context, as in the laboratory experiment, but making it as much like some class of actual behavior setting as possible" [83]. Therefore, experimental tasks in which users access information items that are retrieved from the web in real time are considered. This increases the realism of the experimental procedure, fostering the user engagement with the experimental task.

Addressing RQ3 is done using a *mixed approach*. RQ3 aims at investigating real-time generation of implicit physiological measures for human-information interaction. First, the usefulness of the implicit measures from physiological signals is validated using a controlled laboratory experiment, using a constraint dataset. This allows to better provide measures of the impact that the implicit measures have on the interaction. Then, experimental stimulation is used in order to investigate real-time generation of implicit measures using realistic human-information interaction tasks. This is carried out by implementing a fully integrated information system that uses real-time implicit interaction measures from physiological signals.

To generate implicit measures from physiological signals, the research relies on *machine learning approaches*. Supervised learning schemes are used, which refer to models that are trained based on a set of observations for which "ground truth" measures are available, in order to predict unknown observations [72]. The users' feedback is collected to generate the "ground truth" labels, as the focus is on the users' subjective perception of the information items. Throughout the thesis, different machine learning algorithms are used, which are determined to best suit the needs of the learning models for the specific problem at hand. To elaborate, multiple kernel learning support vector machines [114] is used for RQ1, random forests [18] for RQ2, and regularized linear discriminant analysis [46] for RQ3. The learning algorithms are selected using a deterministic approach, aimed at better suiting the characteristics of the learning problems in each situation such as feature set size, sample size, and class imbalance. It is not the drive of this thesis to develop novel learning algorithms for optimizing physiology-based classification, and therefore a thorough evaluation of which are the best learning approaches for each of the different setups considered throughout addressing RQ1-RQ3 is out of scope. Further details on the rationale for choosing the learning algorithms, as well as implementation details can be found in the respective publications appended to the thesis.

3.3 Research Ethics

While the research presented in this thesis involves human beings, no physical intervention is required with respect to the subjects, and they are not exposed to major security risks. Prior to their participation in the experimental studies, users sign informed consent, and are always informed on their right to withdraw from the experiment at any moment without any adverse consequence. All experimental studies presented in this thesis have been granted ethical approval from the University of Helsinki Ethical review board in humanities and social and behavioral sciences. Special ethical emphasis is placed upon the fact of a) monitoring of users' physiological signals, and b) their application for implicit interaction paradigms. Each of these have separate ethical concerns, which have been carefully taken into account in the design of the research and are discussed below.

Recording physiological signals raise ethical concerns especially regarding the fact that the signals can reflect the inner state of the user and, in many cases, can't be faked or controlled. Additionally, even very simple measures such as EDA have been used to reflect medical conditions [97]. It is thus of the utmost importance that users are aware of which signals are being recorded, and agree with it. Participants involved in the different experiments are further informed that researchers involved in this work have no training in detecting illness-related physiological irregularities, and that the equipment used is not medically certified and thus invalid to provide clinical information. In addition, all physiological data collected from the participants is always kept anonymized. In case participants withdraw from the study, any recorded data is deleted within the following 24 hours.

Implicit interaction paradigms have strong ethical implications on their own as, by definition, they imply users interacting with the systems in a covert fashion. While this might not pose a risk to the users, it still may result in a threat to their privacy, and the ultimate use of implicit signals should always be known by the user. For instance, the user should be aware of the possible use of these implicit signals by third parties, if any. In the presented research, implicit interaction is used with the mere purpose of improving user experience and system performance. Participants are informed that no data is to be used outside of research purposes, nor is to be handed over to third parties.

Chapter 4

Physiology for Implicit Inference of Relevance

Cognitive scientists have long been interested in mapping basic cognitive functions that are highly related to perceiving relevance (e.g., recognition and memory recall [101, 135]) and reacting to relevant stimuli (e.g. implementing intentions [76]). Thus, perception of relevance is a complex cognitive construct, and it is still not completely understood how it happens in the brain, and how it is revealed physiologically [85].

This chapter reports on two experiments that together provide an answer to RQ1: *Can physiological signals be used for implicit relevance inference of textual information?* The experiments were designed in a laboratory setup, under controlled experimental conditions, that allowed better isolation of the relevance correlates. Textual stimuli were presented to participants one at a time, who were asked to assess their relevance to a certain topic. This simplification was introduced to isolate the physiological responses to judging relevance from confounding factors that might occur in scenarios where multiple items are presented simultaneously. Physiological responses to textual content were collected, and predictive models were built to infer relevance metrics from the physiological signals. Further, physiological signals were analyzed in order to locate where and at which point in time relevance decisions are reflected in the physiology.

The first experimental study, summarized in Section 4.1 and reported in Publication I, targets the most basic form of text which are single terms, and evaluates brain activity as the physiological signal to infer relevance. Following, the experimental study which is summarized in Section 4.2 and reported in Publication II, investigates longer pieces of textual content using other physiological signals that have less restrictive time-constraints such as skin conductance and facial muscle activity. Refer to the full publications which are reproduced at the end of this thesis for further details on data collection, processing, and analysis.

4.1 Brain Signals for Relevance Feedback

4.1.1 Overview of the Study

We recorded the EEG signals of forty participants when they assessed relevance in response to term stimuli shown on a screen. The term stimuli were associated with a predetermined topic. Each participant judged relevance for six terms in six topics, for a total of 36 trials. The terms were randomly drawn from a pool of relevant (for each topic) and irrelevant (for all topics) terms defined by experts. We used a balanced setup, i.e., for each topic three relevant and three irrelevant terms were shown to the participant. We randomized the order of the topics and the terms across participants. In addition, the relevance-key assignment was balanced (right or left hand used) between blocks of 12 trials to avoid possibly confounding hemispheric effects. The recorded data reflects the user's subjective relevance judgment of the items (i.e., if a participant assessed the a-priori irrelevant "Morse code" relevant for the topic "Iraq war", it was considered as relevant), as this is the user's real assignment and the corresponding effect is what we would also expect to predict from the brain signals.

4.1.2 Classification Findings

We extracted a set of features from the brain signals, grouped in seven feature sets (e.g., those related to alpha activity, beta activity, event-related potentials, etc.). We then built several predictive models using all, or subsets of these, to evaluate the predictive power of brain signals for relevance feedback. We used data from 38 valid participants, and applied multiple kernel learning (MKL) support vector machines [114] as the learning method to learn classification models from a given set of relevance observations. We used a leave-one-participant-out classification setup, in which for each participant, we trained a model using the data belonging to the rest of the participants, and predicted the unseen trials from the participant. The classifiers using all seven EEG feature sets predicted relevant and irrelevant terms for an unseen participant significantly better than the random baseline, and achieved a mean classification accuracy of 0.54, which represents a mean improvement of 8.30% with respect to the random baseline (see Figure 4.1). The best performing set of features were those capturing ERP and Alpha activity, for which mean classification accuracy raised to

0.56, representing a 11.72% improvement over the baseline. Significance was tested using t-test against random classification accuracy (i.e., 0.5 as we used a balanced setup).

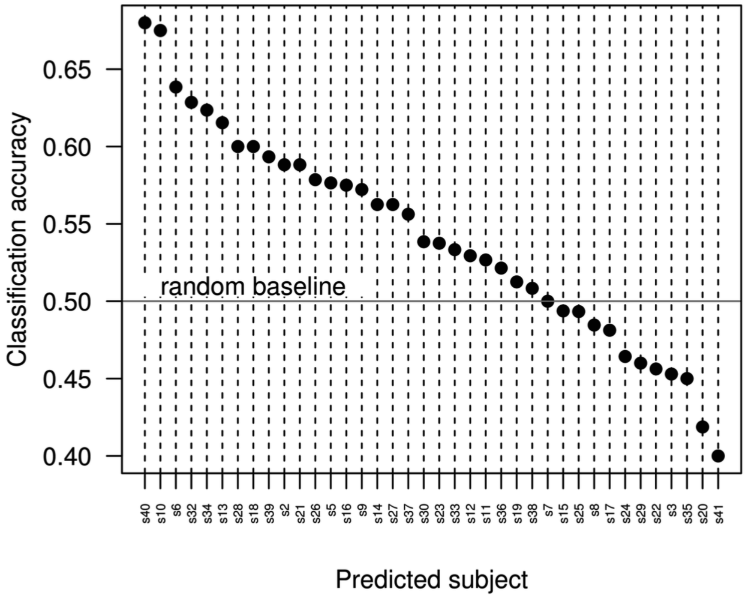


Figure 4.1: Individual classification accuracy for each of the 38 participants, in decreasing order. For each individual participant, models were trained on the data of the remaining participants.

4.1.3 Physiological Findings

The feature sets that were found most effective for the classification were those related to brain activity in the alpha frequency range (8-12 Hz), and those involving the event-related potentials (ERPs). Therefore, these components of the brain signals were further investigated to evaluate the physiological correlates of relevance.

For the alpha activity, we attempted to localize the intra-cranial source of the Alpha using exact low resolution electromagnetic tomography (eLORETA, [93]), and used a F-ratio test to find a maximally significant source localization. Maximal differences between relevant and irrelevant trials were all located in the left frontal lobe, specifically in Brodmann Area 10, which has previously been related to recognition [101], memory retrieval [103], and the evaluation of working memory [135].

To investigate which components of the ERP contributed most to the classification model, we analyzed the mean difference between relevant and irrelevant terms by computing the average relevant and irrelevant ERPs for each participant. The main significant areas were observed in the channels located on the central and parietal regions, with the peak difference in Pz beginning at 477 ms and peaking at 757 ms (see Figure 4.2). The latency and topography of the potential suggest the involvement of a P3-like potential. The high latency and parietal topography coincide with the P3b, thus suggesting that relevance does not affect an early change in orientation, but a later, memory-related effect [98].

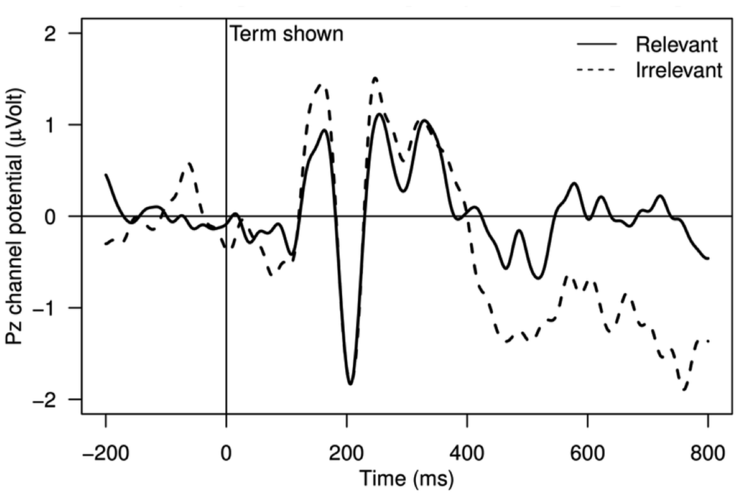


Figure 4.2: Event-related potential (ERP) signal averaged across all 38 participants and trials, locked to term onset. ERP signal shows the significant difference between relevant and irrelevant terms after 450 milliseconds, maximizing at 747 milliseconds.

4.2 Peripheral Physiology for Relevance Feedback

4.2.1 Overview of the Study

We recorded the EDA and the corrugator supercilii activity (CSA) of forty participants while they were examining information items returned by a real information retrieval system. Information items were abstract snippets (40 first words) as extracted from a scientific article database. Participants could type a query of their interest, and the system would return, one at a time and in a randomized order, six abstract snippets: three relevant, and three irrelevant (i.e., balanced setup). The relevant results were retrieved directly based on the ranking model. The irrelevant results were selected randomly with an additional Boolean constraint to exclude results that contained words from the participant's query. The participants then read the abstract snippet and were asked to rate the relevance of the article on a scale from 1 to 10. Participants were instructed to provide the relevance feedback as soon as they made a decision on the relevance, without the need to read the text until the end. Each participant judged the relevance of six snippets in six topics, for a total of 36 trials.

4.2.2 Classification Findings

We extracted a set of features grouped in two feature sets, one for EDA and one for CSA. Features were defined within a time window locked to the moment participants provided relevance feedback. The window spanned from 2 seconds before the explicit rating, which was considered to be sufficient to cover the time between the psychological decision and the physical event (Sternberg stimulus-response model [115]); and 6 seconds following it, as skin conductance response (SCR) can take up to 6 or 7 seconds to reach its peak [36].

We used the data of 36 valid participants, half of which (N=18) were used for feature selection, and the other half for classification evaluation. We used MKL support vector machines [114] as the learning method. We built three classification models, with increasing number of features (top five best features, top ten best features, and all features), and evaluated their performance using a leave-one-participant-out classification setup. Relevance judgments were binarized by considering ratings < 4 irrelevant and ratings > 7 relevant. This led to a slightly unbalanced setup, so we established balanced data to evaluate our models by randomly sampling the relevant and irrelevant sets defined by the smaller set, and repeated the procedure five times. Figure 4.3 shows the individual participant classification accuracies using the top ten best features for each signal, which was the best performing model. The model predicted relevant and irrelevant terms for an unseen participant significantly better than the random baseline, and achieved a mean classification accuracy of 0.57, which represents a mean improvement of 14.22% with respect to the random baseline. Significance was tested using t-test against random classification accuracy (i.e., 0.5 as we used a balanced setup).

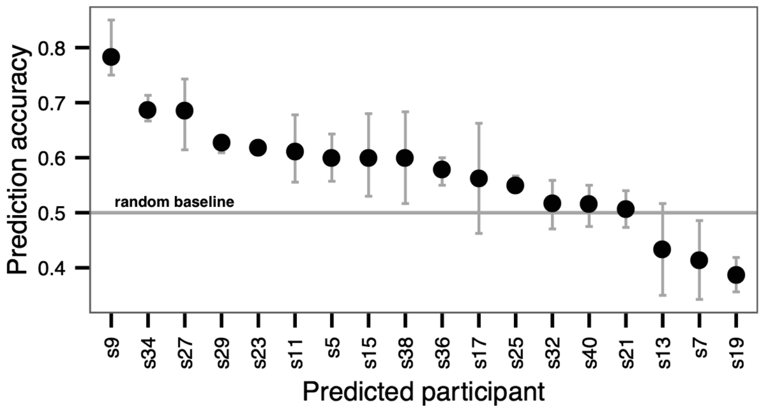


Figure 4.3: Individual classification accuracy for each of the 18 participants, in decreasing order. For each individual participant, models were trained on the data of the remaining participants. Points indicate the mean accuracy, the error bars show the bootstrap confidence intervals.

4.2.3 Physiological Findings

To analyze the EDA and CSA signals within the time window of interest, we computed the grand average across participants and trials of the relevant and irrelevant text snippets. Figure 4.4 shows the grand average based on the mean values with 95% confidence intervals for both signals. In the case of EDA, a difference was visible between relevant and irrelevant trials around 4 to 6 seconds after the explicit relevance judgment. In the case of CSA, a difference was visible around 1 second after the explicit relevance judgment.

We aggregated features from relevant and irrelevant snippets within each participant, with both the mean and the median, and computed repeated-measures analysis of variance (ANOVA). For both signals EDA and CSA, a main effect of time was found, which indicates that the physiological signal changes reliably due to the relevance judgment. However, the direction of the judgment was only significant for electrodermal activity, as indicated by the significant main effect of the relevance. This means that decision-related physiological changes in CSA are either not related to perceived relevance, too weakly related to become visible, not stable enough across time to cause an interactive effect, or not the same between participants.

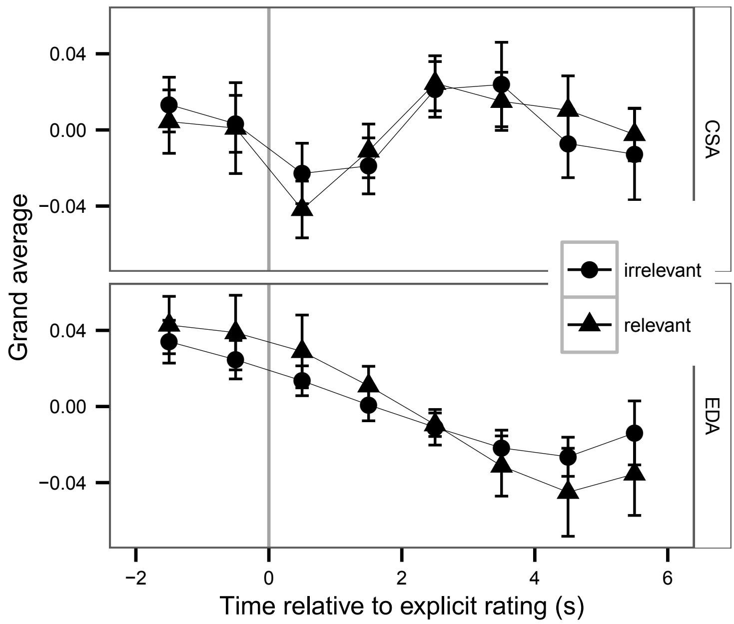


Figure 4.4: Grand average with 95% confidence interval within the 8 second window of the electrodermal activity (EDA, bottom) and corrugator supercilii activity (CSA, top) signals averaged over participants and trials. The vertical gray line at "0" indicates the explicit rating event.

4.3 Contribution

The aim of this chapter is to address RQ1: *Can physiological signals be used for implicit relevance inference of textual information?* The results presented in this chapter demonstrate that it is possible to use physiological signals to infer the user's perceived relevance of textual information. Together, the experimental studies reported in Publications I and II provide fundamental understanding on how to use physiological signals as potential candidates to generate implicit measures of relevance of textual information items, by addressing both single terms and more complex textual information.

First, results show that predictive models based on different physiological sources are able to predict relevance judgments outperforming chance levels. These results provide evidence that physiological responses collected in relevance-assessment tasks of textual information can be used to infer implicit measures of the relevance decision.

Further, the point in time at which the signals are most strongly associated with the relevance judgment, in relation to the relevance decision moment, are located. Results also provide insight on which signals, and which of their components, are the most informative for inferring relevance decisions of users in response to textual information.

Chapter 5

Physiology for Implicit Inference of Affect

Affect refers to the feeling or expression of emotions, and thus is a broad subject. Instead of providing a complete coverage of all possible applications of physiological signals for implicit inference of affect of textual information, which would be out of scope of this thesis, a subset of affective reactions are addressed as the subject of study, focusing the study on non-controlled textual information.

Two divergent scenarios in which users interact with textual information are presented as showcases to provide an answer to RQ2: *Can physiological signals be used for implicit affect inference of non-controlled textual information?* This chapter reports on two experiments that consider affective responses of users when consuming non-controlled information items. The studies were designed such that the users browsed content that was retrieved in real time from the web, and their physiological responses were measured using light-weight physiological sensors. This maximizes ecological validity of the studies to real applications where users access specific web services using different devices, while their responses are measured using wearable technologies. While the experiments were designed in a desktop environment and the signals were recorded using a high-quality physiological amplifier, the studies allow demonstrating the potential for physiological signals to indicate the user's affective responses using light-weight physiological sensors in realistic setups.

The study setups were inspired by widely used existing services that could be enhanced by automatic detection of affective reactions from physiological signals. The first experimental study presented in this chapter, summarized in Section 5.1 and reported in Publication III, investigates affective reactions to news reading in the context of sentiment analysis.

The study was inspired by the most read Italian daily newspaper (both on the Internet and on paper)[1] which asks users to provide affective feedback on the consumed content to mine the overall mood of their readers based on the breaking news. Then, the experimental study summarized in Section 5.2 which is reported in Publication IV, investigates humor detection from comic strips in the context of user-generated content systems. The study was inspired by one of the most used user-generated humorous websites[2], which employs user feedback on the humorousness of the content to modulate the saliency of the extensive amount of data available. Refer to the full publications which are reproduced at the end of this thesis for further details on data collection, processing, and analysis.

5.1 Physiology for Affective Responses to News Articles

5.1.1 Overview of the Study

We designed an experiment to study the physiological correlates of the affective responses of users when consuming textual information encountered in everyday situations. We selected news reading as the experimental stimulus as it has intrinsically associated emotional content, while keeping the user engaged and interested in it. Participants could freely read the news of their choice from their favorite online news sources.

We recorded the EDA signals of 24 participants. Participants were instructed to browse news sites of their interest (typically between 1 and 3) for a minimum of 45 minutes, up to 60 minutes. For each article read, participants were instructed to provide affective feedback according to how they felt after reading the article (i.e, “happy”, “sad”, “angry”, or “neutral”). The goal of the experiment was to find out whether several affective categories could be reliably associated with implicit measures inferred from the physiological signals.

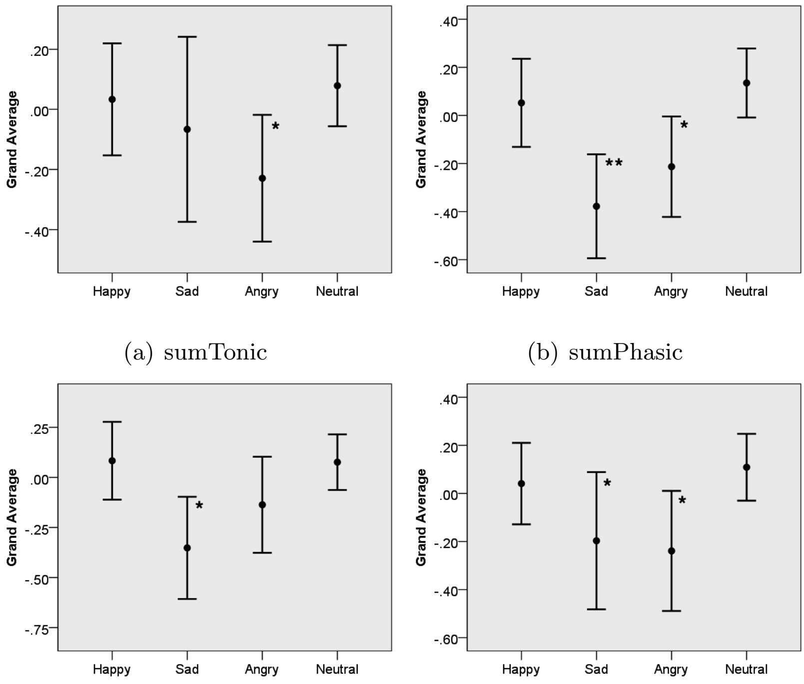
5.1.2 Findings

In order to study how physiological responses were associated with the participants perceived affective content of the news articles, we compared physiological responses to articles where the participant felt “happy”, “sad”, or “angry” against articles where the participant felt “neutral”. The EDA

[1] `Corriere.it`
[2] `9gag.com`

data from 19 valid participants was separated into *phasic*, and *tonic* components, through continuous deconvolution analysis [13], and skin conductance responses (SCRs) were extracted from the signal. We then computed four features from the EDA signal, characterizing the overall tonic activity (*sumTonic*), phasic activity (*sumPhasic*), amplitude of the SCRs (*meanAmpSCR*), and frequency of SCRs (*nSCR*).

We compared the EDA responses of the participants for news articles that received different affective feedback using Linear Mixed Models (LMM) analysis owing to the unbalanced nature of the data. Figure 5.1 shows the aggregated features over participants for each of the affective feedback categories, using the arithmetic mean, with 95% confidence intervals. Asterisks indicate statistical difference with respect to the "neutral" category.



(a) sumTonic (b) sumPhasic

(c) meanAmpSCR (d) nSCR

Figure 5.1: Grand average with 95% confidence intervals for each of the four EDA features and affective feedback categories, with signals averaged over participants and articles. Asterisks indicate statistical difference respect to the "neutral" category, at the levels of $\alpha = .05$ (*), and $\alpha = .001$ (**).

The results indicate that for all of the EDA features, there was no statistical difference between the "happy" and "neutral" categories. However, each EDA feature showed a statistically significant decrease for "sad" or "angry" or both. Thus, news articles that elicited negative affect presented lower levels of skin conductance response, hence, were less arousing.

5.2 Physiology for Affective Responses to Comic Strips

5.2.1 Overview of the Study

We designed an experiment to study physiological signals for implicit inference of affective reactions of users in response to non-controlled textual information. We selected humor as a use case affective reaction, as the ability to infer implicit measures of humor can be applied to a wide range of applications, including real-time adaptive conversational agents, or human-centered implicit tagging. While humor has been investigated using physiological measures, it has mostly been tackled using brain imaging techniques such as functional magnetic resonance imaging (fMRI), which is less practical for realistic HCI applications [5, 129]. On the other hand, it presents the advantage that it has widely been studied using other methods such as content analysis or facial expressions analysis, providing a great ground for comparison.

We recorded EDA, ECG, and EEG from 25 participants while they browsed humorous web comics which were fetched in real-time from a popular user-generated humorous website[3]. Participants were asked to read comic strips for a minimum of 45 minutes up to 60 minutes. Upon reading a strip, participants were instructed to provide affective feedback to indicate whether they found the comic strip funny or not, triggering the presentation of the next strip.

5.2.2 Findings

The goal was to build predictive models from the recorded physiological data to predict whether a comic strip was perceived as funny or not. This was modeled as a two-class prediction setup: "funny" vs. "not funny". We generated a series of features to characterize the changes in each of the physiological signals over the time participants spent reading a given comic strip (i.e., a trial). Features were capturing different signal components at different points of the trial (e.g., gamma EEG activity at the beginning of

[3] `9gag.com`

the trial, number of SCRs at the end of trial, inter-beat-interval over the whole trial, etc.).

The data from 22 valid participants were divided into two subsets of 11 participants each. One subset was used to make decisions about the prediction models, such as determining the best prediction and feature-selection algorithms. The other subset was used to evaluate the predictive models. We used random forests [18] as our prediction algorithm, which was the algorithm that performed best among several algorithms that were tested. Evaluation was carried out using two different approaches: a) *across*, which followed a leave-one-participant-out cross-validation setup; and b) *within*, which followed a leave-one-trial-out cross-validation setup. This allowed to compare prediction performances of models trained using a participant's own data or using others' data. Due to the data imbalance, we used area under the ROC curve (AUC) as a performance metric of the predictive models. AUC is a widely used and sensible measure, even under the class imbalances of our scenario, and it is a comprehensive measure for comparison between the different prediction models.

Figure 5.2 shows the classification performance of the models trained using the different physiological sources in combination and separately, for both evaluation approaches. Results show that combining the three different physiological sources improved the overall performance of the classifiers, and that individually trained models (*within*, average classification performance of 0.73 in terms of AUC) performed better than across participant models (*across*, average classification performance of 0.72 in terms of AUC). Regarding classification models using individual physiological signals, ECG performed worse than EDA and EEG. The results for the latter two in comparison are less clear, as their performances were generally similar. In addition, we built models that used features generated from video recordings of the participants' faces. We evaluated both models using only these features, as well as in combination with physiological features. As expected, models using facial video-recordings outperformed the ones using physiological sources alone. However, on average and especially for some users, physiological signals in combination with video recordings lead to the greatest improvements in classification performances.

5.3 Contribution

The aim of this chapter is to address RQ2: *Can physiological signals be used for implicit affect inference of non-controlled textual information?* The results reported in this chapter demonstrate the use of physiological signals to

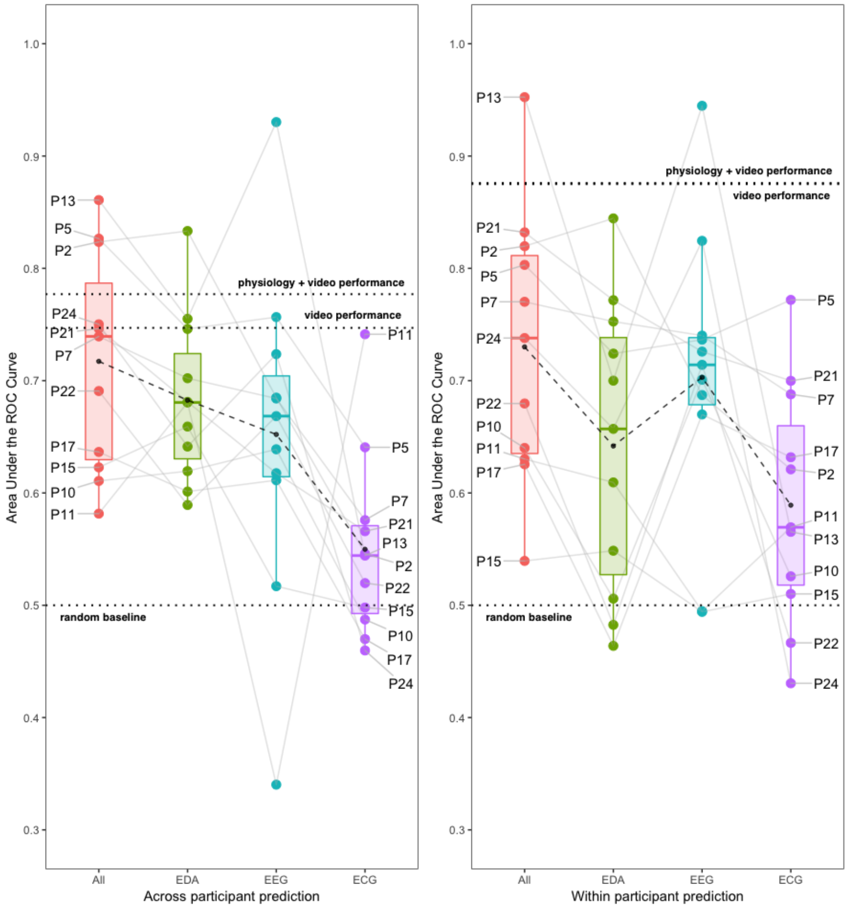


Figure 5.2: Individual classification performance in terms of area under the ROC curve (AUC) for each of the 11 participants and physiological signals (alone, and combined). Performance was computed training the models based on the data of the remaining participants (left), and based on the participants' own data (right). Smaller black dots and dashed lines indicate mean classification performance. Dashed horizontal lines indicate chance performance, as well as averaged classification performance using video-facial recordings alone and in combination with physiological signals.

infer the user's affective responses to textual information items. Together, the experimental studies reported in Publications III and IV, provide fundamental insights on physiological signals as potential candidates to generate implicit measures of affect of textual information fetched in real time from the web.

First, results show that light-weight physiological recording setups can effectively be used to indicate affective responses to non-controlled information content. These results are encouraging when targeting light-weight - potentially wearable - physiological recording technologies that can continuously monitor the users' physiological signals in order to infer their affective responses to textual information.

Further, results show that physiological measures can be used in combination with or as an alternative to other affect recognition methods. Signals such as facial recordings from camera feeds are already ubiquitously available, and the results provide insight on how to use physiological signals in combination with or as an alternative to these other sources of implicit measures of affect.

Chapter 6

Physiology for Real-Time Implicit Interaction

Building computational models to use physiological signals to infer properties of the information items being consumed is essential towards the validation of these signals for real-time implicit interaction. Up to this point, the thesis has addressed physiological signals to generate measures of relevance (RQ1, Chapter 4) and affect (RQ2, Chapter 5) of textual information. The final goal of this thesis is to validate physiological signals not only to infer implicit measures describing the users' responses to textual information, but also to be used for real-time implicit interaction with textual information. Applying these models to working systems that make predictions in real time, adds an additional complexity layer that needs to be investigated. Real systems in real situations involve an intrinsic loss of control over confounding factors as compared to experimental studies. Thus, interaction paradigms that enable gathering implicit measures from physiological signals without disrupting the users' natural behavior, need to be investigated.

In this chapter, brain signals are investigated to generate relevance measures of textual information in real time, and are used as implicit inputs in real information retrieval systems. Out of the approaches investigated when addressing RQ1 and RQ2, relevance prediction from brain signals was selected to exemplify the use of physiology for real-time implicit interaction with textual information for two main reasons. On one hand, perception of relevance is one of the most noteworthy underlying cognitive processes when users interact with information items. On the other hand, brain signals remain one of the most intriguing physiological sources as they directly target the source where relevance judgments are happening, and it is less clear how methods involving brain signals developed under controlled experimental conditions transfer to realistic setups.

This chapter reports on two information retrieval systems that make use of implicit relevance measures on textual information items from brain signals. The systems allow showcasing both the benefits of using implicit measures inferred from physiological signals for information filtering, as well as generating real-time implicit measures; providing an answer to RQ3: *Can physiological signals reliably be used for real-time implicit interaction with textual information?*

The first system, summarized in Section 6.1 and reported in Publication V, uses an approach focused on demonstrating the feasibility of recommending new information based on measures from brain signals only. The users inspect Wikipedia articles while relevance measures of words are extracted from the brain signals. The trained models are then used to recommend new information to the users. The second system presented in this chapter, which is summarized in Section 6.2 and reported in Publication VI, is a fully integrated information retrieval system which uses relevance measures generated in real time based on brain signals and eye movements. The implementation of the system provides an instance of how implicit measures can be generated in real time without disrupting the users' natural behavior. Refer to the full publications which are reproduced at the end of this thesis for further details on system implementation, data collection, processing, and analysis.

6.1 Physiological Signals for Information Filtering

6.1.1 Overview of the System

We implemented a system that closes the human-machine interaction loop. That is, users react to information items and their physiological signals are recorded in order to infer measures of relevance, which are then used to filter information, providing new potentially interesting information to the users. The system uses relevance predictions made on simple words which are then used to infer relevance over all words in the system. Words that are both relevant and informative within the corpus are used to predict the user's search intention and, consequently, to recommend meaningful documents.

6.1.2 Overview of the Study

The document corpus used was the English Wikipedia, and documents were defined as the first six sentences of a Wikipedia article. We recorded the EEG signals of 15 participants while they performed a set of eight reading tasks. In each reading task, participants were first presented with two topics, which were randomly chosen from a list of 30 candidate topics, and were asked to pick one as the topic of interest. Every reading task comprised six trials, each consisting of one sentence from the relevant and one sentence from the irrelevant document. Each trial consisted of the sequential presentation of words, which the participants should "just read" and during which their brain signals were monitored. Then, the participants were presented with the words read previously and asked to rate them as being "relevant" or "irrelevant", which was needed to collect labels for the data analysis.

6.1.3 Relevance Prediction Findings

Participant-specific prediction models were computed for each participant, and the performance was evaluated on a leave-one-reading-task-out cross-validation setup. The classification function was computed with regularized linear discriminant analysis [46, 15], whereby the shrinkage parameter was calculated with an analytic method [77, 109]. We used the AUC to quantify the performance of the classifiers. Figure 6.1 shows the classification performance in terms of AUC for each participant. For 13 out of the 15 participants, the term-relevance prediction models performed significantly better than a prediction model learned based on randomized feedback (hence having AUC > 0.5; $p < 0.05$, within-participant permutation tests with 1000 iterations). For two participants, the predictions were essentially random (AUC not significantly better than 0.5) and they were excluded from the rest of the analyses. These findings are consistent with the ones presented in Section 4.1.2.

6.1.4 Document Recommendation Findings

We used the relevant words –predicted from brain signals– for document retrieval and recommendation, and evaluated the performance in terms of cumulative information gain [62]. The cumulative information gain was defined as the sum of the relevance scores assigned by three experts for the documents that were ranked in the top-30 documents by the retrieval system in response to the brain-based input. Figure 6.2 shows the document

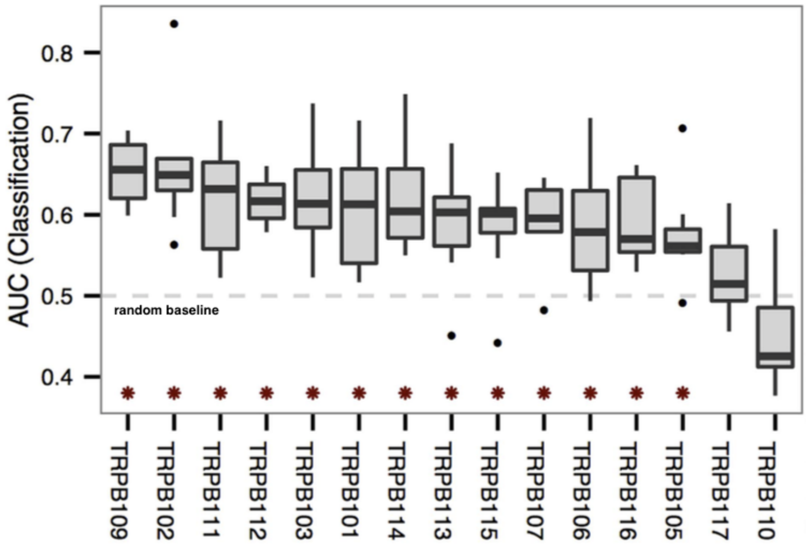


Figure 6.1: Individual classification performances in terms of area under the ROC curve (AUC) of the 15 participants as distributed over the different reading tasks. Asterisks indicate significantly better classification results that the random baseline at the level of $p < 0.05$.

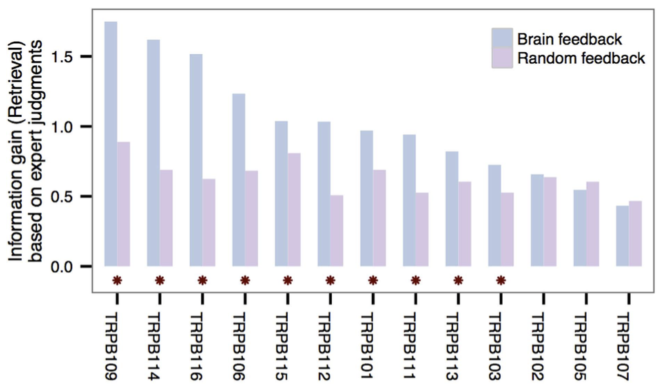


Figure 6.2: Individual retrieval performance for each of the 13 participants in terms of average cumulative information gain (on a scale 0–3) based on the top 30 retrieved documents for the participant. Asterisks indicate a significantly better pooled information gain based on brain feedback than random feedback retrieval, at the level of of $p < 0.05$.

retrieval performance for each of the 13 participants that presented relevance predictions above chance levels, in terms of mean information gain (on a 3-point likert scale). For each participant, the figure shows the mean information gain over all reading tasks based on brain feedback (blue bars) and randomized feedback (purple bars). For 10 participants, the brain feedback resulted in significantly greater information gain as compared to randomized feedback ($p < 0.05$; two-sided Wilcoxon test).

6.2 Physiological Signals for Real-Time Implicit Interaction

6.2.1 Overview of the System

In addition to showing the usefulness of physiological signals to infer implicit measures to provide adaptation, such as information filtering, another main challenge is to generate the measures in real time, while users interact with the information system. To demonstrate real-time implicit interaction, we implemented a fully integrated system that supports real-time implicit relevance feedback from physiological signals. The implicit relevance feedback, together with explicit relevance feedback, is fed into the user model, which allows the interactive information retrieval engine to refine accordingly the next iteration of results to show to the user.

Before the user is able to engage in using the fully integrated system, the physiological classifier, which is in charge of generating real-time implicit relevance feedback, needs to be trained. This user-specific "calibration phase" involves collecting enough user data in order to train the personalized classifier, which is a common procedure in BCI systems [87]. To collect training data for the physiological classifier, we use a dataset that closely matches the application domain and consists of a set of topics and associated keywords. During the "calibration phase" users are prompted with a list of five topics, randomly selected from the dataset, and are asked to select one as the topic of interest. Then, a series of keywords are shown to the user, who is asked later to indicate the relevance to the selected topic. This procedure is repeated for several topics until the system has gathered enough data to train the physiological classifier. This procedure is similar to the one described in Section 4.1.

Once the classifier has been trained, the user can proceed to the search task that benefits from real-time implicit relevance feedback based on the physiological signals (i.e., the "online phase"). The user initiates the search by entering a query, which results in the first set of results retrieved by the system. To direct the search, users can open a view that displays a set of keywords that are potentially relevant to their search. The users can examine these keywords and provide explicit relevance feedback on one of the keywords (by clicking on it to indicate that the keyword is relevant). While users examine the keywords, the physiological classifier generates implicit relevance feedback for each keyword fixated upon. The system then takes into account both the explicit relevance feedback, and the implicit feedback to estimate the user's search intentions and return a new set of results.

6.2.2 Overview of the Study

Sixteen participants took part in the final evaluation experiment. The goal was to evaluate the feasibility to generate real-time implicit measures from physiological signals, while users performed a simulated-work task interacting with a fully integrated information system. After engaging in the "calibration phase", participants were given a topic and asked to interact with the system in order to gather enough information to write an essay about it. For evaluation purposes, the participants were prompted at the end of each iteration with a dialog asking them to label the relevance of the keywords they had fixated on.

6.2.3 Offline Relevance Prediction Findings

In order to evaluate the feasibility and performance of the system in predicting relevance, we first evaluated the classification performance in the "calibration phase". The data used in the "calibration phase" was controlled and had the advantage that the same dataset was used to train the classification models of the different users. Implicit relevance measures were inferred from the EEG and eye tracking signals, for each keyword on the screen that was fixated upon. For this purpose, high-dimensional feature vectors were extracted from short epochs of the EEG signals, and the fixation durations were appended to the corresponding feature vectors. The classification function was computed with regularized linear discriminant analysis [46, 15], whereby the shrinkage parameter was calculated with an analytic method [77, 109]. Further details regarding the physiological classifier, in terms of feature extraction, signal contribution to the overall model, and performance, can be found in the appended Publication VI, as well as in the original publication by Wenzel et al. (2017) [130].

Classification performance was computed in terms of AUC and was evaluated using a $10x10$ cross-validation setup. To quantify the significance and the effect sizes of the implicit relevance feedback from the physiological signals, we compared the classification performances against performances from prediction models learned from randomized feedback. Standard permutation tests were applied for significance testing [51].

Figure 6.3 presents the individual classification performances in the "calibration phase". Results show that the classification performance significantly outperformed random predictions for 13 participants (hence having AUC > 0.5; $p < 0.05$, within-participant permutation tests with 1000 iterations). For three participants, the predictions were essentially random (AUC not significantly better than 0.5). These findings are consistent with the ones presented in Section 4.1.2, and Section 6.1.3.

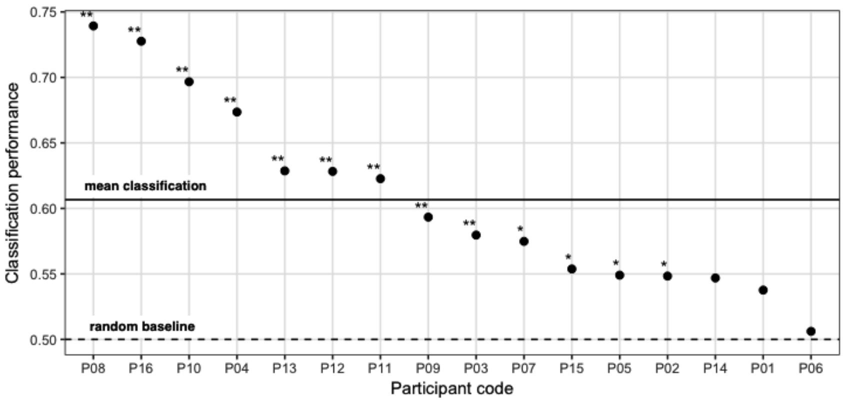


Figure 6.3: Individual classification performances of the 16 participants in terms of area under the ROC curve (AUC) in the "calibration phase". Asterisks indicate significantly better classification results that the random baseline at the levels of $p < 0.05$ (*), and $p < 0.001$ (**).

6.2.4 Real-Time Relevance Prediction Findings

The main challenge is to assess how well the classification performance achieved in the controlled "calibration phase" transfers to the "online phase". In this phase, the implicit relevance measures are inferred from physiological signals in real time, while users are engaged in a realistic information-seeking task rather than a relevance assessment task. Further, the data presented to the users is also generated in real time, as opposed to the controlled dataset used in the "calibration phase".

We implemented a user model that can estimate both the user's current search intent and the certainty of the relevance feedback provided by the user based on the limited feedback from keywords. The explicit feedback is considered very certain, while the physiology-based implicit feedback is uncertain, so their certainty is inferred from data. For example, if the implicit feedback is in line with the previous history of feedback, particularly with explicit feedback, then it will be inferred as certain and will contribute to the user model. However, if it contradicts the system's current belief, then the inference system will consider it as an outlier. Further details on the implementation of the user intent model can be found in the appended Publication VI.

The participants whose classification performance in the "calibration phase" was not significantly better than random were discarded from further analyses. Furthermore, a participant had to be rejected from the analysis due to problems in the logging during the execution of the experimental task. For the remaining twelve participants, we studied how well the classification performance achieved using the controlled "calibration phase" transferred to the realistic search task in the "online phase". This was done by computing the classification performances in terms of AUC both as directly predicted from the classifier in real-time, and as predicted by the user intent model.

Figure 6.4 shows the classification performances in the offline "calibration phase" and for the real-time search task in the "online phase". Overall, results show that the classification performances achieved using the controlled dataset in the "calibration phase" transferred to the real search task in the "online phase". The overall distribution across participants remained above random classification levels when considering the relevance predictions based only on the physiological signals (Figure 6.4, middle). Furthermore, the user model was able to correct the noisy implicit relevance feedback, improving the overall performance and leading to performances above chance levels for 10 of the 12 participants (Figure 6.4, right).

6.3 Contribution

The aim of this chapter is to address RQ3: *Can physiological signals reliably be used for real-time implicit interaction with textual information?* The results reported in this chapter demonstrate the use of physiological signals in real time to generate measures to be used for implicit interaction in realistic setups. Together, the systems reported in Publications V and VI provide evidence of how physiological signals can be implemented

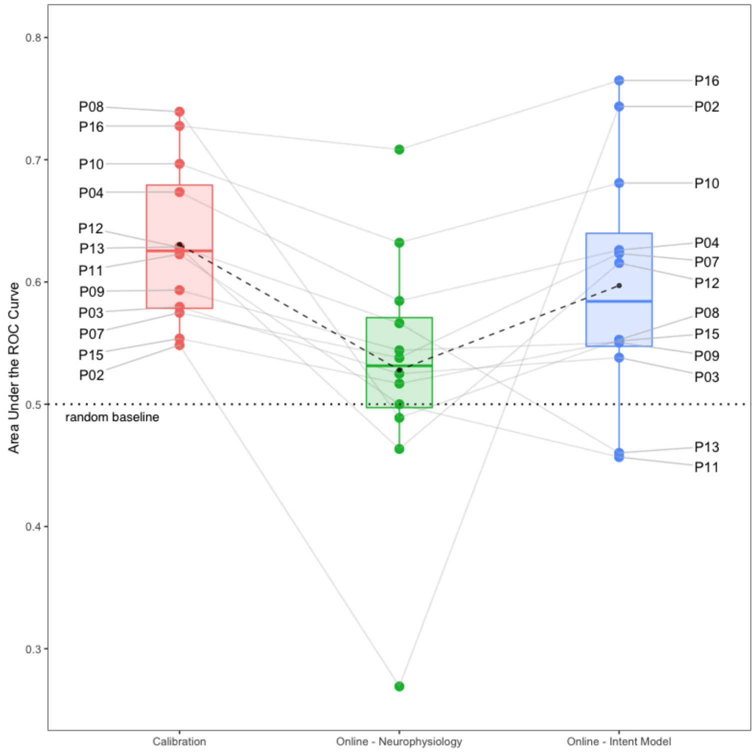


Figure 6.4: Individual classification performance in terms of area under the ROC curve (AUC) for each of the 12 participants. Left: offline prediction in the “calibration phase”. Middle: physiological prediction in the “online phase”. Right: user intent model prediction in the “online phase”. Smaller black dots and dashed lines indicate mean classification performance.

in interactive information systems in order to provide implicit interaction channels for real-time adaptation.

First, results show that implicit measures extracted from physiological signals on textual information can be used to recommend new information to users. The results indicate that physiological signals can be used not only to extract implicit measures of relevance, but also that these can effectively be used for system adaptation, closing the bio-cybernetic loop.

Further, results show the feasibility of reliably generating implicit physiological measures in real time, while users perform realistic tasks using interactive information systems. Models that are trained using a separate relevance assessment task prove to be transferable to realistic simulated-work tasks, providing reliable implicit relevance measures in real time.

Chapter 7

Discussion

This thesis has explored the use of physiological signals for implicit interaction with textual information. The research was broken down into three research questions (RQ1 - RQ3), which were addressed in Chapters 4-6 respectively. Below I provide a summary of the main contributions before discussing the implications of the research. Then, I discuss the limitations of the proposed approach and point out directions for continuing the research, prior to concluding remarks.

7.1 Summary of the Main Contributions

Chapter 4 studied the potential of physiological signals to infer relevance properties of textual information items (RQ1). Following, Chapter 5 addressed physiological signals in regard to implicit affect inference of non-controlled textual information (RQ2). Finally, Chapter 6 tackled physiological signals for real-time implicit interaction with textual information (RQ3). The research presented in these chapters led to the following main contributions which, together, present a comprehensive approach to physiological signals for implicit interaction with textual information.

- C1: Demonstrating the use of physiological signals to infer the user's perceived relevance of textual information (RQ1)
  - C1.1: Building predictive models from physiological signals that outperform chance levels
  - C1.2: Identifying the spatial and temporal structure of relevance reflected in different physiological signals

The work summarized in Chapter 4, reported in Publications I and II, addresses RQ1 by demonstrating the use of different physiological signals to infer measures of relevance of textual information items. In Publication I, we use brain signals as measured by EEG to infer relevance of single terms. Then, in Publication II, we use skin conductance and facial muscle activity, as measured by EDA and fEMG respectively, to infer relevance of text snippets. We demonstrate in both cases that physiology-based predictive models outperform random models, representing classification improvements of 11.72% for selected features of the brain-based models, and 14.22% for selected features of the EDA and fEMG-based models. While the accuracies of the predictions are to some extent low, the studies establish physiological signals as valid candidates to generate implicit relevance feedback of textual information. In addition, we identify how and at which point in time relevance judgments are most prominently reflected in the physiology. More specifically, relevance of terms from brain signals are shown to be most reflected in the Alpha activity (8-12 Hz), and on the event-related potentials, maximizing at 757 milliseconds after stimulus onset. Relevance of text snippets is reflected with a peak in the EDA signals following 4 to 6 seconds after the relevance judgments, while most discriminative differences are found in fEMG around 1 second after the relevance judgment. The contribution to the body of work does not only rely on the fact that physiological signals are studied in response to textual information, but also on the insights given regarding the temporal and spatial structure of relevance as reflected in the physiological signals.

C2: Demonstrating the use of physiological signals to infer the user's affective responses to textual information (RQ2)

- C2.1: Using light-weight physiological recordings to indicate affective responses to non-controlled information content
- C2.2: Proving that physiological measures can be used in combination or as alternative to common affect recognition methods such as facial recordings

The work summarized in Chapter 5, reported in Publications III and IV, addresses RQ2 by demonstrating various uses for physiological signals to indicate affective responses to information items. The work especially targets measuring physiological responses to information items which are fetched in real time from the web, and that consist mostly of textual content. In Publication III we target skin conductance, as measured by EDA, to indicate affective responses to news reading. We demonstrate that a light-weight physiological recording sensor can be used to indicate measures of affective responses to information content being consumed under realistic situations. We show that lower levels of skin conductance are found when reading news that elicit negative emotions, as compared to neutral or positive news articles. Then, in Publication IV we target light-weight sensors to measure cardiovascular activity, skin conductance, and brain activity (using ECG, EDA, and EEG respectively) to capture the user's affective reactions to humorous content. We showcase the potential for physiological signals to generate meaningful implicit measures to capture humor appraisal, achieving averaged classification performances of 0.73 in terms of AUC when using all of the recorded physiological signals in user-specific trained models. Moreover, the highest classification performances are found when combining physiological signals with video recordings of the participant's face (with an average classification performance of 0.88 in terms of AUC). This illustrates the potential for physiological sources to be used as a complement or as an alternative to other well-established implicit measures of affect detection such as facial recordings. Thereafter, together, the results provide a contribution to the previous research work as they demonstrate that light-weight physiological signals can successfully be used to indicate affective responses of textual information, without the need of having highly controlled measurement environments.

C3: Demonstrating the use of physiological signals for real-time implicit interaction with textual information (RQ3)

- C3.1: Recommending new information using implicit measures extracted from physiological signals
- C3.2: Reliably generating implicit physiological measures in real time

The work summarized in Chapter 6, reported in Publications V and VI, addresses RQ3 by demonstrating the use of physiological signals for real-time implicit interaction with textual information. In Publication V, we use measures of relevance extracted from brain signals in order to recommend new information to the users. The publication validates the use of implicit measures from physiological signals for implicit interaction, constituting presumably one of the first proof-of-concept systems that have performed automatic information filtering on the basis of brain activity alone. In fact, the recommended information leads to an improved information gain for 75% of the participants for which relevance predictions are statistically better than by chance. Then, in Publication VI, we demonstrate the generation of real-time implicit measures from physiological signals. We present a fully integrated information retrieval system that uses real-time implicit relevance measures from brain signals and eye movements, in combination with explicit feedback, to estimate the user intent and present the next iteration of results. The system exhibits an interaction paradigm that uses a separate task to calibrate the physiological classifier, which allows the generation of reliable implicit physiological measures in real time for 80% of the participants for which the relevance predictions are statistically better than by chance in the separate calibration task. Jointly, these results represent a substantial contribution to the body of work, as, to the extent of my knowledge, they are the first to demonstrate textual information filtering on the basis of physiological signals alone, as well as the first to demonstrate the generation of implicit relevance measures on textual information in real time within a fully integrated system.

7.2 Implications of the Research

Overall, the research holds implications for general human-computer interaction (HCI), as C1-C3 contribute to emerging interaction approaches between man and machines. Implicit interaction paradigms are increasingly being investigated as they provide novel interaction possibilities that do not rely on the users' active engagement with the systems. These interactive paradigms are raising attention as the interaction with computerized systems and devices is becoming more pervasive and ubiquitous. While physiological recordings become less intrusive and wearable, they provide a completely new perspective and exciting possibilities for implicit interactive scenarios in HCI.

The main contributions of this thesis can be grouped roughly into two research themes: a) capturing the user's responses to information items (C1 and C2), and b) providing implicit input for human interaction with information items (C3). The two research themes hold separate implications for a variety of research fields.

The first theme, which concerns using physiological signals to characterize the user's responses to information items, holds more direct implications for fields related to information access, namely information retrieval, recommender systems, and user modeling. C1 provides insight on how physiological signals can be used to gather implicit relevance measures of textual content. Perception of relevance is crucial for capturing the user's information needs, which is a central aspect for improved retrieval, recommendation and personalization of information. Further, C1 and C2 have implications for automatic content annotation, as physiological signals present themselves as candidates to generate additional meta-data to better describe the information items within a domain. In Publication III we present the physiological text annotation framework, which exemplifies how the different findings presented in this thesis can be used for implicit content annotation. The framework proposes attaching physiological measures to textual content as additional meta-data. These annotations can then be used in addition to, or complementing, other annotation mechanisms that could be explicit (e.g., user tags), or implicit (e.g., dwell time). One of the possible uses for the annotations is to enable information recommendation. Affective recommender systems have been defined as recommender systems that use affective responses of the users as contextual information to provide recommendations [121]. While the gathering of affective measures have been centered using explicit feedback or affective signals such as facial recordings [92], C2 provides insight on how these affective measures can be gathered using physiological signals.

The second theme, which concerns using physiological signals to provide implicit input for human interaction with information items, is especially relevant for the field of physiological computing systems. Physiological computing systems refer to systems that use physiological signals for real-time adaptation [40]. C3 bridges the fields of information retrieval and physiological computing systems, as it demonstrates the use of physiological signals for real-time adaptation using a fully integrated information retrieval system. More specifically, C3 holds great implications for brain-computer interfaces (BCIs), namely for passive brain-computer interfaces. Traditional BCIs use active engagement of the users in order to provide control input to computerized systems. Passive BCIs, on the contrary, focus on monitoring the user's brain signals for implicit interaction, that is, for BCIs that do not require active engagement from the user [30, 134]. C3 provides solutions for passive BCIs in information retrieval, contributing to the field of neuro-information retrieval systems [53].

7.3 Limitations and Future Work

This thesis has provided evidence that it is possible to generate reliable implicit measures describing the interaction of users with textual information, besting chance levels. However, one of the main limitations encountered in the research is the generally low classification performances, namely in the relevance case as the recognition rates are consistently below 70%. Thereafter, when using these types of measures for implicit interaction, it is necessary to investigate ways to make use of these high-throughput, but uncertain measures. One possibility is to build a user model that considers previous relevance feedback values (which could potentially come from different sources), and uses it to estimate the likelihood of a classification to be accurate. This is the approach presented in Section 6.2. Other possibilities include using descriptive information about the information domain to mitigate the downsides of misclassified measures, or collecting repeated measures in order to accumulate evidence towards more reliable predictions. An example is the work by Wenzel et al. (2017), where they use brain and eye signals to infer the relevance of keywords known to belong to a certain category, in order to predict the category of interest (e.g., "chair" and "table" belonging to the category "furniture"). The study shows the benefits of using domain knowledge by repeatedly aggregating individual predictions on words in order to infer broader categories of interest [130].

Another limitation that I consistently encountered at various stages of my research is the high variability across users in terms of classification performances. More importantly, be it studying measures of relevance or affect, results show that predictive methods do not perform better than the chance level for a non-negligible portion of the participants. This can be attributed to the fact that brain-computer interface control has proven to not work for around 15 - 30% of users [2, 16]. Analogously, EDA has also proven to be unresponsive in the case of some users [125]. Novel methods have been proposed to mitigate the issue [126], however, it remains unclear how to completely avoid this drawback. Future work addressing physiological signals for implicit interaction should take this into consideration by, for instance, involving richer models that could bridge this limitation.

The recording devices used in the research, while being considered non intrusive in terms of physical interventions, are still cumbersome (especially EEG recordings), which represent a limitation regarding the applicability of the proposed methods to widely used applications. This is especially problematic as these sensors take time to set up, and in some cases a trained person is required to set them up correctly. Novel wearable devices that allow the measurement of EEG and other physiological signals in a less cumbersome manner already exist [47, 56]. However, at this stage, it is still unclear how these devices can be used to implement the type of interaction paradigms proposed in this thesis, especially due to the increase in signal-to-noise ratio, which would most likely worsen the reliability of the implicit measures extracted from physiological signals. As a matter of fact, with this concern in mind, light-weight physiological recording setups were considered in the studies presented in Section 5. These setups were investigated to generate measures of affect, and it is less clear how they would work for other types of implicit measures. Thus, future work should study in more detail what type of implicit measures describing the interaction between users and information items, and under which setups, can be inferred using lighter-weight, and ultimately wearable recording technology [73, 80, 133].

Section 6.2 presented an interaction paradigm that facilitates the generation of real-time implicit measures from physiological signals, without disrupting the user's natural behavior when performing search tasks. The main hindrance of the proposed approach is that the user must spend some amount of time prior to using the system, in order to provide calibration data to train the user-specific models. This presents a limitation towards the application of these methods for everyday use, as users are unlikely to be willing to spend large amounts of time to set up the sensors and calibrate the system before they are able to engage in their task. To mitigate this issue, future work should investigate alternative approaches to unobtrusively collect training data for providing real-time implicit interaction. Models that can learn from the interaction of other users [42, 43], or self-training methods [95, 110] are some of the research directions to be investigated.

Lastly, while this thesis has addressed physiological signals to indicate relevance and affect separately, in some cases one cannot completely separate the two constructs. In fact, affect plays a role in the relevance judgment process, as factors such as the overall search experience, system responsiveness or difficulty of the task may influence the user's affective state and, in turn, the way in which the information items are perceived. Such an affective component of the relevance judgment process has been defined in information science as affective relevance [108]. For instance, a user might feel annoyed by a system that does not match her expectations, in turn influencing the way she perceives the results of the system. In this case, the affective state of the user plays a role in the process of perceiving the relevance of information items [32]. Future work research should extend the presented results, assessing scenarios where relevance and affect might be more intertwined, further helping to bring the notion of affective relevance from information science theory to real information systems [10].

7.4 Conclusion

This thesis is the result of close to four years of research that has led to six original publications. The thesis goes beyond the individual publications, by providing an understanding on how to utilize physiological signals as implicit inputs for human interaction with textual information. This is achieved by focusing on three research areas. First, physiological signals are investigated to indicate measures of how relevant users perceive textual information. Second, instances of how physiological signals can be used to indicate affective reactions to textual information are provided. Last, real-time implicit interaction using physiological signals is demonstrated by presenting two systems that allow information filtering based on physiological signals alone, as well as real-time generation of implicit measures from physiological signals.

The thesis provides both basic and applied research on the topic of implicit interaction with textual information using physiological signals. This sheds light on the potential that these signals have to enrich everyday human-computer interaction. Furthermore, it helps to identify the main challenges that remain to be addressed for such interactive setups to be adopted in everyday systems. Altogether, this thesis contributes to the development of the next generation computer systems that will utilize ubiquitously available physiological measurements through wearable recordings, as implicit interactive inputs.

References

[1] J. Aasman, G. Mulder, and L. J. Mulder. Operator effort and the measurement of heart-rate variability. *Human factors*, 29(2):161–170, 1987.

[2] L. Acqualagna, L. Botrel, C. Vidaurre, A. Kübler, and B. Blankertz. Large-Scale Assessment of a Fully Automatic Co-Adaptive Motor Imagery-Based Brain Computer Interface. *PloS one*, 11(2):e0148886, feb 2016.

[3] J. Allen. Photoplethysmography and its application in clinical physiological measurement. *Physiological Measurement*, 28(3):R1–R39, mar 2007.

[4] M. Ames and M. Naaman. Why we tag. In *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '07*, page 971, New York, New York, USA, 2007. ACM Press.

[5] O. Amir and I. Biederman. The Neural Correlates of Humor Creativity. *Frontiers in Human Neuroscience*, 10(November):597, 2016.

[6] J. L. Andreassi. *Psychophysiology: Human behavior & physiological response*. Psychology Press, 2000.

[7] I. Arapakis, K. Athanasakos, and J. M. Jose. A comparison of general vs personalised affective models for the prediction of topical relevance. In *Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval*, SIGIR '10, pages 371–378, New York, NY, USA, 2010. ACM.

[8] I. Arapakis, I. Konstas, and J. M. Jose. Using facial expressions and peripheral physiological signals as implicit indicators of topical relevance. In *Proceedings of the 17th ACM international conference on Multimedia*, pages 461–470. ACM, 2009.

[9] O. Barral, M. J. Eugster, T. Ruotsalo, M. M. Spapé, I. Kosunen, N. Ravaja, S. Kaski, and G. Jacucci. Exploring peripheral physiology as a predictor of perceived relevance in information retrieval. In *Proceedings of the 20th International Conference on Intelligent User Interfaces*, IUI '15, pages 389–399, New York, NY, USA, 2015. ACM.

[10] O. Barral and G. Jacucci. Applying physiological computing methods to study psychological, affective and motivational relevance. In *Symbiotic Interaction*, pages 35–46. Springer, 2014.

[11] O. Barral, I. Kosunen, and G. Jacucci. No need to laugh out loud: Predicting humor appraisal of comic strips based on physiological signals in a realistic environment. *ACM Transactions on Computer-Human Interaction*, 24(6):40, Dec. 2017.

[12] O. Barral, I. Kosunen, T. Ruotsalo, M. M. Spapé, M. J. A. Eugster, N. Ravaja, S. Kaski, and G. Jacucci. Extracting relevance and affect information from physiological text annotation. *User Modeling and User-Adapted Interaction*, 26(5):493–520, 2016.

[13] M. Benedek and C. Kaernbach. A continuous measure of phasic electrodermal activity. *Journal of Neuroscience Methods*, 190(1):80 – 91, 2010.

[14] S. Bitzer and P. van der Smagt. Learning EMG control of a robotic hand: towards active prostheses. In *Proceedings 2006 IEEE International Conference on Robotics and Automation, 2006. ICRA 2006.*, pages 2819–2823. IEEE, 2006.

[15] B. Blankertz, S. Lemm, M. Treder, S. Haufe, and K.-R. Müller. Single-trial analysis and classification of ERP components – A tutorial. *NeuroImage*, 56(2):814–825, 2011.

[16] B. Blankertz, C. Sannelli, S. Halder, E. M. Hammer, A. Kübler, K.-R. Müller, G. Curio, and T. Dickhaus. Neurophysiological predictor of SMR-based BCI performance. *NeuroImage*, 51(4):1303–1309, 2010.

[17] W. Boucsein. *Electrodermal Activity.* Springer US, Boston, MA, 2nd edition, 2012.

[18] L. Breiman. Random Forests. *Machine Learning*, 45(1):5–32, 2001.

[19] A.-M. Brouwer, M. Hogervorst, B. Reuderink, Y. van der Werf, and J. van Erp. Physiological signals distinguish between reading emotional and non-emotional sections in a novel. *Brain-Computer Interfaces*, 2(2-3):76–89, 2015.

[20] A.-M. Brouwer, M. A. Hogervorst, J. B. F. van Erp, T. Heffelaar, P. H. Zimmerman, and R. Oostenveld. Estimating workload using EEG spectral power and ERPs in the n-back task. *Journal of Neural Engineering*, 9(4):045008, aug 2012.

[21] A.-M. Brouwer, M. A. Neerincx, V. Kallen, L. van der Leer, and M. ten Brinke. Eeg alpha asymmetry, heart rate variability and cortisol in response to virtual reality induced stress. *Journal of Cybertherapy & Rehabilitation*, 4(1):21–34, 2011.

[22] A.-M. Brouwer, N. van Wouwe, C. Mühl, J. van Erp, and A. Toet. Perceiving blocks of emotional pictures and sounds: effects on physiological variables. *Frontiers in human neuroscience*, 7:295, 2013.

[23] A.-M. Brouwer, T. O. Zander, J. B. F. van Erp, J. E. Korteling, and A. W. Bronkhorst. Using neurophysiological signals that reflect cognitive or affective state: six recommendations to avoid common pitfalls. *Frontiers in neuroscience*, 9:136, 2015.

[24] G. Buscher, A. Dengel, R. Biedert, and L. V. Elst. Attentive documents: Eye tracking as implicit feedback for information retrieval and beyond. *ACM Transactions on Interactive Intelligent Systems*, 1(2):1–30, jan 2012.

[25] G. Buscher, A. Dengel, and L. van Elst. Query expansion using gaze-based feedback on the subdocument level. In *Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '08, pages 387–394, New York, NY, USA, 2008. ACM.

[26] J. T. Cacioppo, R. E. Petty, M. E. Losch, and H. S. Kim. Electromyographic activity over facial muscle regions can differentiate the valence and intensity of affective reactions. *Journal of Personality and Social Psychology*, 50(2):260, 1986.

[27] J. T. Cacioppo, L. G. Tassinary, and A. J. Fridlund. The skeletomotor system. In *Principles of psychophysiology: Physical, social, and inferential elements*, pages 325–384. Cambridge University Press, 1990.

[28] J. T. E. Cacioppo, L. G. E. Tassinary, and G. G. Berntson. *Handbook of Psychophysiology.* Cambridge University Press, Cambridge, 3rd edition, 2007.

[29] F. Chan, Yong-Sheng Yang, F. Lam, Yuan-Ting Zhang, and P. Parker. Fuzzy EMG classification for prosthesis control. *IEEE Transactions on Rehabilitation Engineering*, 8(3):305–311, 2000.

[30] E. B. Coffey, A.-M. Brouwer, E. S. Wilschut, and J. B. van Erp. Brain-machine interfaces in space: Using spontaneous rather than intentionally generated brain signals. *Acta Astronautica*, 67(1-2):1–11, 2010.

[31] C. Conati and H. Maclaren. Empirically building and evaluating a probabilistic model of user affect. *User Modeling and User-Adapted Interaction*, 19(3):267–303, 2009.

[32] E. Cosijn and P. Ingwersen. Dimensions of relevance. *Information Processing & Management*, 36(4):533–550, 2000.

[33] B. Cowley, M. Filetti, K. Lukander, J. Torniainen, A. Henelius, L. Ahonen, O. Barral, I. Kosunen, T. Valtonen, M. Huotilainen, N. Ravaja, and G. Jacucci. The psychophysiology primer: A guide to methods and a broad review with a focus on human–computer interaction. *Foundations and Trends® in Human—Computer Interaction*, 9(3-4):151–308, 2016.

[34] F. Daniel, F. Casati, P. Silveira, M. Verga, and M. Nalin. Beyond Health Tracking: A Personal Health and Lifestyle Platform. *IEEE Internet Computing*, 15(4):14–22, jul 2011.

[35] J. Davidson, B. Liebald, J. Liu, P. Nandy, T. Van Vleet, U. Gargi, S. Gupta, Y. He, M. Lambert, B. Livingston, et al. The youtube video recommendation system. In *Proceedings of the fourth ACM conference on Recommender systems*, RecSys '10, pages 293–296. ACM, 2010.

[36] M. E. Dawson, A. M. Schell, D. L. Filion, and G. G. Berntson. The electrodermal system. In J. T. Cacioppo, L. G. Tassinary, and G. Berntson, editors, *Handbook of Psychophysiology*, pages 157–181. Cambridge University Press, 3rd edition, 2007.

[37] M. J. Eugster, T. Ruotsalo, M. M. Spapé, I. Kosunen, O. Barral, N. Ravaja, G. Jacucci, and S. Kaski. Predicting term-relevance from

brain signals. In *Proceedings of the 37th International ACM SIGIR Conference on Research & Development in Information Retrieval*, SIGIR '14, pages 425–434, New York, NY, USA, 2014. ACM.

[38] M. J. A. Eugster, T. Ruotsalo, M. M. Spapé, O. Barral, N. Ravaja, G. Jacucci, and S. Kaski. Natural brain-information interfaces: Recommending information by relevance inferred from human brain signals. *Scientific Reports*, 6(1):38580, dec 2016.

[39] S. Fairclough and K. Gilleade. Construction of the biocybernetic loop. In *Proceedings of the 14th ACM international conference on Multimodal interaction - ICMI '12*, page 571, New York, New York, USA, 2012. ACM Press.

[40] S. H. Fairclough. Fundamentals of physiological computing. *Interacting with computers*, 21(1-2):133–145, Jan. 2009.

[41] S. H. Fairclough and K. Gilleade. *Advances in Physiological Computing*. Springer Publishing Company, Incorporated, 2014.

[42] S. Fazli, C. Grozea, M. Danóczy, B. Blankertz, F. Popescu, and K.-R. Müller. Subject independent eeg-based bci decoding. In Y. Bengio, D. Schuurmans, J. D. Lafferty, C. K. I. Williams, and A. Culotta, editors, *Advances in Neural Information Processing Systems 22*, pages 513–521. Curran Associates, Inc., 2009.

[43] S. Fazli, F. Popescu, M. Danóczy, B. Blankertz, K.-R. Müller, and C. Grozea. Subject-independent mental state classification in single trials. *Neural Networks*, 22(9):1305–1312, nov 2009.

[44] R. A. Fisher. *The design of experiments*. Oliver And Boyd; Edinburgh; London, 1937.

[45] J. Frey. Heart rate monitoring as an easy way to increase engagement in human-agent interaction. In *Proceedings of the 2nd International Conference on Physiological Computing Systems, PhyCS*, pages 129–136. INSTICC, ScitePress, 2015.

[46] J. H. Friedman. Regularized Discriminant Analysis. *Journal of the American Statistical Association*, 84(405):165–175, mar 1989.

[47] G. Gargiulo, R. A. Calvo, P. Bifulco, M. Cesarelli, C. Jin, A. Mohamed, and A. van Schaik. A new EEG recording system for passive dry electrodes. *Clinical Neurophysiology*, 121(5):686–693, 2010.

[48] A. Gevins and M. E. Smith. Neurophysiological measures of cognitive workload during human-computer interaction. *Theoretical Issues in Ergonomics Science*, 4(1-2):113–131, 2003.

[49] J.-E. Golenia, M. A. Wenzel, and B. Blankertz. Live demonstrator of EEG and eye-tracking input for disambiguation of image search results. In *Symbiotic Interaction*, pages 81–86. Springer, 2015.

[50] G. Gonzalez, J. L. de la Rosa, M. Montaner, and S. Delfin. Embedding Emotional Context in Recommender Systems. In *2007 IEEE 23rd International Conference on Data Engineering Workshop*, pages 845–852. IEEE, apr 2007.

[51] P. I. Good. *Permutation tests : a practical guide to resampling methods for testing hypotheses.* Springer, 2nd edition, 2000.

[52] J. Gwizdka. Characterizing relevance with eye-tracking measures. In *Proceedings of the 5th Information Interaction in Context Symposium*, IIiX '14, pages 58–67, New York, NY, USA, 2014. ACM.

[53] J. Gwizdka and J. Mostafa. NeuroIR 2015: SIGIR 2015 Workshop on Neuro-Physiological Methods in IR Research. In *ACM SIGIR Forum*, volume 49, pages 83–88. ACM, jan 2015.

[54] J. Gwizdka and J. Mostafa. NeuroIIR: Challenges in Bringing Neuroscience to Research in Human-Information Interaction. In *Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval - CHIIR '17*, pages 437–438, New York, New York, USA, 2017. ACM Press.

[55] T. Han, X. Xiao, L. Shi, J. Canny, and J. Wang. Balancing Accuracy and Fun: Designing Camera Based Mobile Games for Implicit Heart Rate Monitoring. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15*, pages 847–856, New York, New York, USA, 2015. ACM Press.

[56] C. J. Harland, T. D. Clark, and R. J. Prance. High resolution ambulatory electrocardiographic monitoring using wrist-mounted electric potential sensors. *Measurement Science and Technology*, 14(7):923–928, jul 2003.

[57] D. Harman and Donna. Relevance feedback revisited. In *Proceedings of the 15th annual international ACM SIGIR conference on Research and development in information retrieval - SIGIR '92*, pages 1–10, New York, New York, USA, 1992. ACM Press.

[58] S. Haufe, M. S. Treder, M. F. Gugler, M. Sagebaum, G. Curio, and B. Blankertz. EEG potentials predict upcoming emergency brakings during simulated driving. *Journal of Neural Engineering*, 8(5):056001, oct 2011.

[59] S. V. Ioannou, A. T. Raouzaiou, V. A. Tzouvaras, T. P. Mailis, K. C. Karpouzis, and S. D. Kollias. Emotion recognition through facial expression analysis based on a neurofuzzy network. *Neural Networks*, 18(4):423 – 435, 2005.

[60] G. Jacucci, O. Barral, P. Daee, M. Wenzel, B. Serim, T. Ruotsalo, P. Pluchino, J. Freeman, L. Gamberini, S. Kaski, and B. Blankertz. Integrating neurophysiological relevance feedback in intent modeling for information retrieval. *Submitted to Journal*, 2017.

[61] G. Jacucci, S. Fairclough, and E. T. Solovey. Physiological Computing. *Computer*, 48(10):12–16, oct 2015.

[62] K. Järvelin and J. Kekäläinen. Cumulated gain-based evaluation of IR techniques. *ACM Transactions on Information Systems*, 20(4):422–446, oct 2002.

[63] A. Kapoor, P. Shenoy, and Desney Tan. Combining brain computer interfaces with vision for object categorization. In *2008 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8. IEEE, jun 2008.

[64] J.-P. Kauppi, M. Kandemir, V.-M. Saarinen, L. Hirvenkari, L. Parkkonen, A. Klami, R. Hari, and S. Kaski. Towards brain-activity-controlled information retrieval: Decoding image relevance from MEG signals. *NeuroImage*, 112:288–98, may 2015.

[65] D. Kelly. Methods for Evaluating Interactive Information Retrieval Systems with Users. *Foundations and Trends® in Information Retrieval*, 3(1–2):1–224, 2009.

[66] D. Kelly and X. Fu. Elicitation of term relevance feedback: An investigation of term source and context. In *Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '06, pages 453–460, New York, NY, USA, 2006. ACM.

[67] D. Kelly and J. Teevan. Implicit Feedback for Inferring User Preference: A Bibliography. *ACM SIGIR Forum*, 37(2):18–28, sep 2003.

[68] A. Kobsa and W. Wahlster. *User Models in Dialog Systems.* Springer Berlin Heidelberg, 1989.

[69] S. Koelstra, C. Muhl, M. Soleymani, J.-S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras. Deap: A database for emotion analysis using physiological signals. *IEEE Transactions on Affective Computing*, 3(1):18–31, Jan 2012.

[70] S. Koelstra and I. Patras. Fusion of facial expressions and EEG for implicit affective tagging. *Image and Vision Computing*, 31(2):164–174, 2013.

[71] I. Kosunen, M. Salminen, S. Järvelä, A. Ruonala, N. Ravaja, and G. Jacucci. RelaWorld. In *Proceedings of the 21st International Conference on Intelligent User Interfaces - IUI '16*, pages 208–217, New York, New York, USA, 2016. ACM Press.

[72] S. B. Kotsiantis. Supervised machine learning: A review of classification techniques. *Informatica*, 31:249–268, 2007.

[73] O. E. Krigolson, C. C. Williams, A. Norton, C. D. Hassall, and F. L. Colino. Choosing MUSE: Validation of a Low-Cost, Portable EEG System for ERP Research. *Frontiers in neuroscience*, 11:109, 2017.

[74] D. Krusienski, E. Sellers, D. McFarland, T. Vaughan, and J. Wolpaw. Toward enhanced P300 speller performance. *Journal of Neuroscience Methods*, 167(1):15–21, 2008.

[75] K. Kuikkaniemi, T. Laitinen, M. Turpeinen, T. Saari, I. Kosunen, and N. Ravaja. The influence of implicit and explicit biofeedback in first-person shooter games. In *Proceedings of the 28th international conference on Human factors in computing systems - CHI '10*, page 859, New York, New York, USA, 2010. ACM Press.

[76] A. Lagioia, S. Eliez, M. Schneider, J. S. Simons, M. Van der Linden, and M. Debbané. Neural correlates of reality monitoring during adolescence. *NeuroImage*, 55(3):1393–1400, apr 2011.

[77] O. Ledoit and M. Wolf. A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, 88(2):365–411, feb 2004.

[78] R. Leeb, D. Friedman, G. R. Müller-Putz, R. Scherer, M. Slater, and G. Pfurtscheller. Self-paced (asynchronous) BCI control of a

wheelchair in virtual environments: a case study with a tetraplegic. *Computational intelligence and neuroscience*, 2007:79642, 2007.

[79] M. Lewandowska, J. Rumiński, T. Kocejko, and J. Nowak. Measuring pulse rate with a webcam – a non-contact method for evaluating cardiac activity. In *Proceedings of the Federated Conference on Computer Science and Information Systems*, pages 405–410. IEEE, 2011.

[80] C. L. Lisetti and F. Nasoz. Using noninvasive wearable computers to recognize human emotions from physiological signals. *Eurasip Journal on Applied Signal Processing*, 2004(11):1672–1687, 2004.

[81] I. Lopatovska. Searching for good mood: Examining relationships between search task and mood. *Proceedings of the American Society for Information Science and Technology*, 46(1):1–13, 2009.

[82] I. Lopatovska and I. Arapakis. Theories, methods and current research on emotions in library and information science, information retrieval and human—computer interaction. *Information Processing & Management*, 47(4):575 – 592, 2011.

[83] J. E. McGrath. Methodology matters: Doing research in the behavioral and social sciences. In *Readings in Human–Computer Interaction*, pages 152–169. Elsevier, 2nd edition, 1995.

[84] Y. Moshfeghi and J. M. Jose. An effective implicit relevance feedback technique using affective, physiological and behavioural features. In *Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '13, pages 133–142, New York, NY, USA, 2013. ACM.

[85] Y. Moshfeghi, L. R. Pinto, F. E. Pollick, and J. M. Jose. Understanding relevance: An fmri study. In *Proceedings of the 35th European Conference on IR Research*, ECIR '13, pages 14–25. Springer, 2013.

[86] J. Mostafa and J. Gwizdka. Deepening the role of the user: Neurophysiological evidence as a basis for studying and improving search. In *Proceedings of the 2016 ACM on Conference on Human Information Interaction and Retrieval*, pages 63–70. ACM, 2016.

[87] A. Nijholt, D. Tan, G. Pfurtscheller, C. Brunner, J. d. R. MillÃ¡n, B. Allison, B. Graimann, F. Popescu, B. Blankertz, and K. R. MÃ¼ller. Brain-computer interfacing for intelligent systems. *IEEE Intelligent Systems*, 23(3):72–79, May 2008.

[88] J. J. S. Norton, D. S. Lee, J. W. Lee, W. Lee, O. Kwon, P. Won, S.-Y. Jung, H. Cheng, J.-W. Jeong, A. Akce, S. Umunna, I. Na, Y. H. Kwon, X.-Q. Wang, Z. Liu, U. Paik, Y. Huang, T. Bretl, W.-H. Yeo, and J. A. Rogers. Soft, curved electrode systems capable of integration on the auricle as a persistent brain-computer interface. *Proceedings of the National Academy of Sciences of the United States of America*, 112(13):3920–5, mar 2015.

[89] F. T. Oliveira, A. Aula, and D. M. Russell. Discriminating the relevance of web search results with measures of pupil size. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, pages 2209–2212, New York, NY, USA, 2009. ACM.

[90] J. J. Oresko, Zhanpeng Jin, Jun Cheng, Shimeng Huang, Yuwen Sun, H. Duschl, and A. C. Cheng. A Wearable Smartphone-Based Platform for Real-Time Cardiovascular Disease Detection Via Electrocardiogram Processing. *IEEE Transactions on Information Technology in Biomedicine*, 14(3):734–740, may 2010.

[91] M. Pantic and L. Rothkrantz. Toward an affect-sensitive multimodal human-computer interaction. *Proceedings of the IEEE*, 91(9):1370–1390, Sept 2003.

[92] M. Pantic and A. Vinciarelli. Implicit human-centered tagging. *IEEE Signal Processing Magazine*, 26(6):173–180, nov 2009.

[93] R. D. Pascual-Marqui, C. M. Michel, and D. Lehmann. Low resolution electromagnetic tomography: A new method for localizing electrical activity in the brain. *International Journal of Psychophysiology*, 18(1):49–65, 1994.

[94] P. Perego, A. Moltani, and G. Andreoni. Sport monitoring with smart wearable system. *Studies in health technology and informatics*, 177:224–228, 2012.

[95] K. Pfeuffer, M. Vidal, J. Turner, A. Bulling, and H. Gellersen. Pursuit calibration. In *Proceedings of the 26th annual ACM symposium on User interface software and technology - UIST '13*, pages 261–270, New York, New York, USA, 2013. ACM Press.

[96] R. W. Picard and S. B. Daily. Evaluating affective interactions: Alternatives to asking what users feel. In *CHI Workshop on Evaluating Affective Interfaces: Innovative Approaches*, pages 2119–2122. ACM New York, NY, 2005.

[97] M.-Z. Poh, T. Loddenkemper, C. Reinsberger, N. C. Swenson, S. Goyal, M. C. Sabtala, J. R. Madsen, and R. W. Picard. Convulsive seizure detection using a wrist-worn electrodermal activity and accelerometry biosensor. *Epilepsia*, 53(5):e93–e97, may 2012.

[98] J. Polich. Updating p300: An integrative theory of P3a and P3b. *Clinical Neurophysiology*, 118(10):2128–2148, 2007.

[99] A. T. Pope, E. H. Bogart, and D. S. Bartolome. Biocybernetic system evaluates indices of operator engagement in automated task. *Biological Psychology*, 40(1-2):187–195, may 1995.

[100] L. J. Prinzel, F. G. Freeman, M. W. Scerbo, P. J. Mikulka, and A. T. Pope. A Closed-Loop System for Examining Psychophysiological Measures for Adaptive Task Allocation. *The International Journal of Aviation Psychology*, 10(4):393–410, oct 2000.

[101] C. Ranganath, M. K. Johnson, and M. Desposito. Prefrontal activity associated with working memory and episodic long-term memory. *Neuropsychologia*, 41(3):378 – 389, 2003.

[102] N. Ravaja. Contributions of psychophysiology to media research: Review and recommendations. *Media Psychology*, 6(2):193–235, 2004.

[103] M. D. Rugg, P. C. Fletcher, C. D. Frith, R. S. J. Frackowiak, and R. J. Dolan. Differential activation of the prefrontal cortex in successful and unsuccessful memory retrieval. *Brain*, 119:2073–2083, 1996.

[104] T. Ruotsalo, G. Jacucci, P. Myllymäki, and S. Kaski. Interactive intent modeling. *Communications of the ACM*, 58(1):86–92, dec 2014.

[105] J. Salojärvi, I. Kojo, J. Simola, and S. Kaski. Can relevance be inferred from eye movements in information retrieval. In *Proceedings of the 4th Workshop on Self-Organizing Maps*, WSOM '03, pages 261–266, 2003.

[106] G. Salton and C. Buckley. Improving retrieval performance by relevance feedback. *Readings in information retrieval*, 24(5):355–363, 1997.

[107] T. S. Saponas, D. S. Tan, D. Morris, R. Balakrishnan, J. Turner, and J. A. Landay. Enabling always-available input with muscle-computer interfaces. In *Proceedings of the 22nd annual ACM symposium on User interface software and technology - UIST '09*, page 167, New York, New York, USA, 2009. ACM Press.

[108] T. Saracevic. Relevance reconsidered. In *Proceedings of the second conference on conceptions of library and information science (CoLIS 2)*, pages 201–218. ACM Press, 1996.

[109] J. Schäfer and K. Strimmer. A Shrinkage Approach to Large-Scale Covariance Matrix Estimation and Implications for Functional Genomics. *Statistical Applications in Genetics and Molecular Biology*, 4(1), jan 2005.

[110] F. Schettini, F. Aloise, P. Aricò, S. Salinari, D. Mattia, and F. Cincotti. Self-calibration algorithm in an asynchronous P300-based brain–computer interface. *Journal of Neural Engineering*, 11(3):035004, jun 2014.

[111] C. Setz, B. Arnrich, J. Schumm, R. La Marca, G. Troster, and U. Ehlert. Discriminating Stress From Cognitive Load Using a Wearable EDA Device. *IEEE Transactions on Information Technology in Biomedicine*, 14(2):410–417, mar 2010.

[112] M. Soleymani and M. Pantic. Human-centered implicit tagging: Overview and perspectives. In *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics*, SMC '12, pages 3304–3309, Oct 2012.

[113] M. Soleymani and M. Pantic. Multimedia implicit tagging using eeg signals. In *Proceedings of the IEEE International Conference on Multimedia and Expo*, ICME '13, pages 1–6. IEEE, 2013.

[114] S. Sonnenburg, G. Rätsch, C. Schäfer, and B. Schölkopf. Large Scale Multiple Kernel Learning. *Journal of Machine Learning Research*, 7(Jul):1531–1565, 2006.

[115] S. Sternberg. The discovery of processing stages: Extensions of donders' method. *Acta Psychologica*, 30(0):276 – 315, 1969.

[116] J. Teevan, C. Alvarado, M. S. Ackerman, and D. R. Karger. The perfect search engine is not enough. In *Proceedings of the 2004 conference on Human factors in computing systems - CHI '04*, pages 415–422, New York, New York, USA, 2004. ACM Press.

[117] D. J. Terbizan, B. A. Dolezal, and C. Albano. Validity of Seven Commercially Available Heart Rate Monitors. *Measurement in Physical Education and Exercise Science*, 6(4):243–247, dec 2002.

[118] Y. Tian, T. Galery, G. Dulcinati, E. Molimpakis, and C. Sun. Facebook emotions: Reactions and Emojis. In *Proceedings of the Fifth International Workshop on Natural Language Processing for Social Media*, SocialNLP'17, pages 11–16. Association for Computational Linguistics, 2017.

[119] M. Tkalčič, U. Burnik, and A. Košir. Using affective parameters in a content-based recommender system for images. *User Modeling and User-Adapted Interaction*, 20(4):279–311, 2010.

[120] M. Tkalcic, U. Burnik, and A. KoÅ¡ir. Using affective parameters in a content-based recommender system for images. *User Modeling and User-Adapted Interaction*, 20(4):279–311, 2010.

[121] M. Tkalcic, A. Kosir, and J. Tasic. Affective recommender systems: the role of emotions in recommender systems. In *Proceedings of The RecSys 2011 Workshop on Human Decision Making in Recommender Systems*, pages 9–13. Citeseer, 2011.

[122] A. Van Boxtel and M. Jessurun. Amplitude and bilateral coherency of facial and jaw-elevator emg activity as an index of effort during a two-choice serial reaction task. *Psychophysiology*, 30(6):589–604, 1993.

[123] R. van Lien, M. Neijts, G. Willemsen, and E. J. C. de Geus. Ambulatory measurement of the ECG T-wave amplitude. *Psychophysiology*, 52(2):225–237, feb 2015.

[124] I. Veldhuizen, A. Gaillard, and J. De Vries. The influence of mental fatigue on facial emg activity during a simulated workday. *Biological Psychology*, 63(1):59–78, 2003.

[125] P. Venables and D. Mitchell. The effects of age, sex and time of testing on skin conductance activity. *Biological Psychology*, 43(2):87 – 101, 1996.

[126] C. Vidaurre and B. Blankertz. Towards a Cure for BCI Illiteracy. *Brain Topography*, 23(2):194–198, jun 2010.

[127] A. Vinciarelli, N. Suditu, and M. Pantic. Implicit Human-Centered Tagging. In *Proceedings of the IEEE International Conference on Multimedia and Expo*, ICME '09, pages 1428–1431. IEEE, jun 2009.

[128] W. Waterink and A. Van Boxtel. Facial and jaw-elevator emg activity in relation to changes in performance level during a sustained information processing task. *Biological Psychology*, 37(3):183–198, 1994.

[129] K. K. Watson, B. J. Matthews, and J. M. Allman. Brain activation during sight gags and language-dependent humor. *Cerebral Cortex*, 17(2):314–324, 2007.

[130] M. A. Wenzel, M. Bogojeski, and B. Blankertz. Real-time inference of word relevance from electroencephalogram and eye gaze. *Journal of Neural Engineering*, 14(5):056007, oct 2017.

[131] R. W. White, I. Ruthven, and J. M. Jose. A study of factors affecting the utility of implicit relevance feedback. In *Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval - SIGIR '05*, page 35, New York, New York, USA, 2005. ACM Press.

[132] G. F. Wilson. An Analysis of Mental Workload in Pilots During Flight Using Multiple Psychophysiological Measures. *The International Journal of Aviation Psychology*, 12(1):3–18, jan 2002.

[133] S. W. Wong, R. H. Chan, and J. N. Mak. Spectral modulation of frontal EEG during motor skill acquisition: A mobile EEG study. *International Journal of Psychophysiology*, 91(1):16–21, 2014.

[134] T. O. Zander and C. Kothe. Towards passive brain–computer interfaces: applying brain–computer interface technology to human–machine systems in general. *Journal of Neural Engineering*, 8(2):025005, apr 2011.

[135] J. X. Zhang, H.-C. Leung, and M. K. Johnson. Frontal activations associated with accessing and evaluating information in working memory: an fmri study. *NeuroImage*, 20(3):1531 – 1539, 2003.

[136] X. Zhang, X. Chen, W.-h. Wang, J.-h. Yang, V. Lantz, and K.-q. Wang. Hand gesture recognition and virtual game control based on 3D accelerometer and EMG sensors. In *Proceedingsc of the 13th international conference on Intelligent user interfaces - IUI '09*, page 401, New York, New York, USA, 2008. ACM Press.