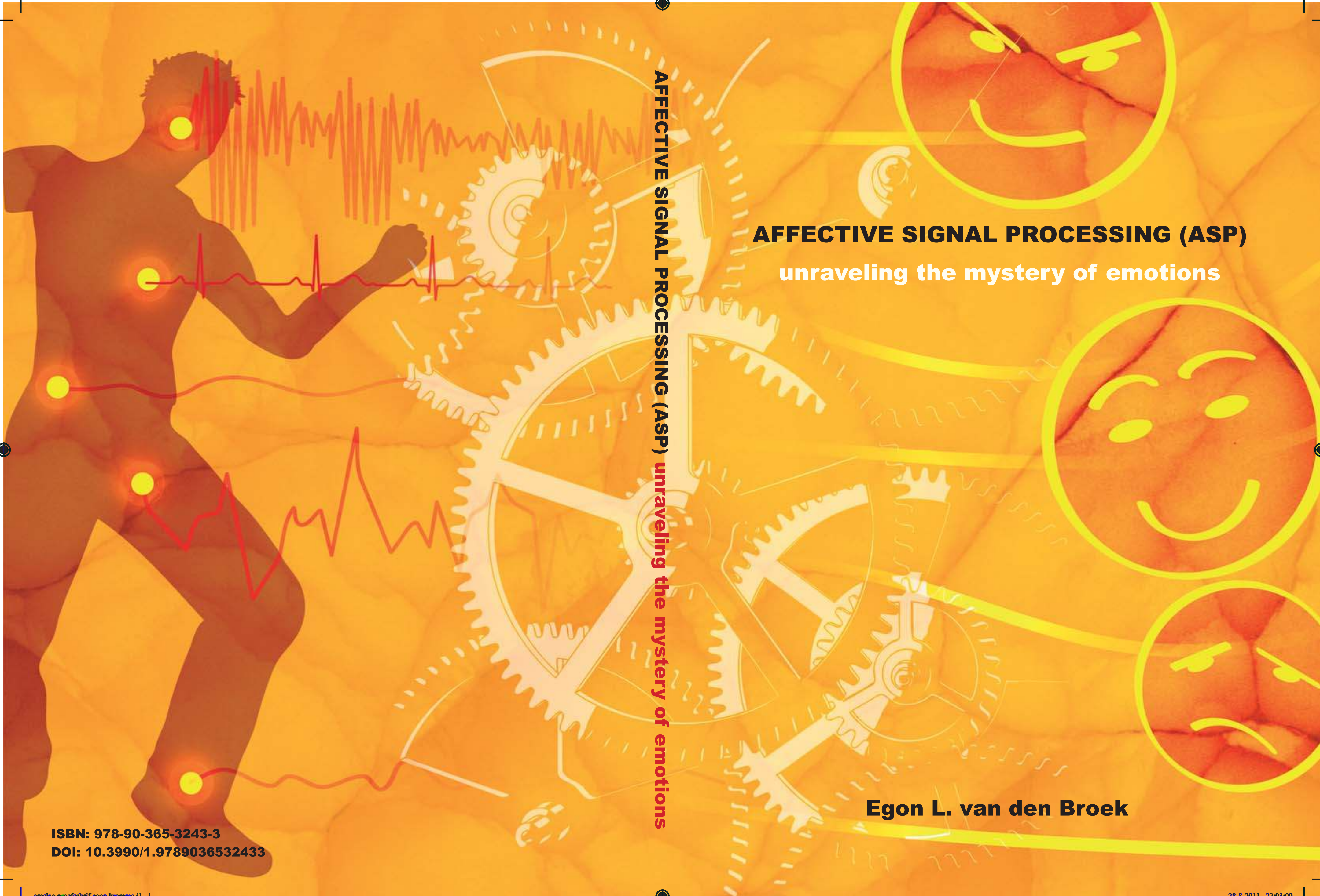


AFFECTIVE SIGNAL PROCESSING (ASP)

unraveling the mystery of emotions

Egon L. van den Broek

AFFECTIVE SIGNAL PROCESSING (ASP) unraveling the mystery of emotions

ISBN: 978-90-365-3243-3
DOI: 10.3990/1.9789036532433

Stellingen

behorende bij het proefschrift

Affective Signal Processing (ASP):
Unraveling the mystery of emotions

1. ASP zal op termijn middelen verschaffen om mensen te manipuleren (maar misschien vinden ze dat niet eens erg).
2. ASP lijdt onder het uitblijven van standaarden.
3. Een voldoende voorwaarde om de kansverdeling van een continue stochast X op een oneindig interval te karakteriseren door haar centrale momenten (bekend als het Hamburger momentenprobleem) is:
$$\sum_{n=1}^{\infty} \left(\inf_{i \leq n} (E[X^{2i}])^{\frac{1}{2i}} \right)^{-1} = \infty.$$
Daaraan is voldaan door biosignalen die een Laplace of normale verdeling hebben, hetgeen meestal het geval is. Hierbij zijn X's centrale momenten gedefinieerd als:
$$E[(X - \bar{x})^n] = \int_{-\infty}^{+\infty} (x - \overline{x})^n f_X(x) dx,$$
waarbij $\bar{x}$ de gemiddelde waarde van X is en f_X de dichtheidsfunctie van X. Het eindige rijtje van de eerste n centrale momenten (bijvoorbeeld $n = 4$) is een compacte representatie van X en geeft, als zodanig, een alternatief voor andere signaal decompositie technieken (bijvoorbeeld Fourier en wavelets), wat ook interessant is voor ASP, vanuit zowel affectief en computationeel oogpunt.
4. "Als je kunt meten waarover je spreekt en je het uit kunt drukken in getallen dan weet je er iets over." (William Thomson, beter bekend als Lord Kelvin, 1824–1907, 1883). Toch is het, om cognitieve engineering (zoals ASP) van theorie naar de praktijk te brengen, nodig om ook van onduidelijke modellen gebruik te maken.
5. Nu de samenleving ICT omarmt, worden ethische kwesties in verband met ICT belangrijker. Helaas worstelt de ethiek nog met het veroveren van een plaats binnen de techniek.
6. Multidisciplinair onderzoek is nog geen interdisciplinair onderzoek. In het eerste geval is vaak nog sprake van onbegrip voor elkaars methoden, theorieën en cultuur; in het tweede geval zijn deze problemen grotendeels opgelost.
7. Onderwijs is nog steeds het ondergeschoven kindje van de Nederlandse universiteiten.

Egon L. van den Broek
Wenen, Oostenrijk, 1 augustus 2011

Propositions

belonging to the Ph.D.-thesis

Affective Signal Processing (ASP):
Unraveling the mystery of emotions

1. ASP will eventually provide the means to manipulate people (but, perhaps they won't even mind).

2. ASP suffers from a lack of standardization.

3. A sufficient condition for the probability distribution of a continuous random variable X to be characterized by its central moments (i.e., the Hamburger moment problem) for an infinite interval is given by:

$$\sum_{n=1}^{\infty} \left(\inf_{i \leq n} (E[X^{2i}])^{\frac{1}{2i}} \right)^{-1} = \infty,$$

   which holds for biosignals that have a Laplace or normal distribution, as is usually the case. With X's central moments being defined as:

$$E[(X - \bar{x})^n] = \int_{-\infty}^{+\infty} (x - \overline{x})^n f_X(x) dx,$$

   where $\bar{x}$ is the average value of X and f_X is the density function of X. The finite series of the first n central moments (e.g., $n = 4$) is a compact representation of X and provides, as such, an alternative to other signal decomposition techniques (e.g., Fourier and wavelets), which is also interesting for ASP, from both an affective and a computational point of view.

4. "*... when you can measure what you are speaking about, and express it in numbers, you know something about it ...*" (William Thomson; a.k.a. Lord Kelvin, 1824–1907, 1883). Although true, to bring cognitive engineering (e.g., ASP) from theory to practice, ill defined models must also be embraced.

5. With society embracing ICT, ethical issues in relation to ICT are increasing in importance. Regrettably, they are still struggling to find their way into engineering.

6. Multidisciplinary research is not the same as interdisciplinary research. With the first, incomprehension for each other's methods, theories, and culture is often still present; with the latter, these problems have largely been resolved.

7. Education is still the red headed stepchild of the Dutch universities.

Egon L. van den Broek
Vienna, Austria, August 1, 2011

Affective Signal Processing (ASP)

Unraveling the mystery of emotions

Egon L. van den Broek

Ph.D. dissertation committee:

Chairman and Secretary
prof. dr. M. C. Elwenspoek, University of Twente, The Netherlands
Promotores:
prof. dr. ir. A. Nijholt, University of Twente, The Netherlands
prof. dr. T. Dijkstra, Radboud University Nijmegen, The Netherlands
Assistent-promotor:
dr. J. H. D. M. Westerink, Philips Research, The Netherlands
Members:
prof. dr. P. M. G. Apers, University of Twente, The Netherlands
prof. dr. A. Esposito, Second University of Naples, Italy
International Institute for Advanced Scientific Studies, Italy
prof. dr. ir. H. J. Hermens, University of Twente, The Netherlands /
Roessingh Research and Development, The Netherlands
prof. dr. ir. E. Hoenkamp, Queensland University of Technology, Australia
prof. dr. L. R. B. Schomaker, University of Groningen, The Netherlands

Paranimfen:
Joris H. Janssen, M.Sc., Eindhoven University of Technology, The Netherlands /
Philips Research, The Netherlands
Frans van der Sluis, M.Sc., University of Twente, The Netherlands /
Radboud University Medical Center Nijmegen, The Netherlands

CTIT Ph.D.-thesis series No. 11-204 (ISSN: 1381-3617)
Centre for Telematics and Information Technology (CTIT)
P.O. Box 217, 7500 AE Enschede, The Netherlands

SIKS Dissertation series No. 2011-30
The research reported in this thesis has been carried out under the auspices of SIKS, the Dutch Research School for Information and Knowledge Systems.

This book was typeset by the author using LaTeX 2$_\varepsilon$.
Cover: Design and graphics by Wilson Design, Uden, The Netherlands.
Printing: Ipskamp Drukkers, Enschede, The Netherlands.

AFFECTIVE SIGNAL PROCESSING (ASP)

UNRAVELING THE MYSTERY OF EMOTIONS

PROEFSCHRIFT

ter verkrijging van
de graad doctor aan de Universiteit Twente,
op gezag van de rector magnificus,
prof. dr. H. Brinksma,
volgens besluit van het College voor Promoties
in het openbaar te verdedigen
op vrijdag 16 september 2011 om 14.45 uur

door

Egidius Leon van den Broek

geboren op 22 augustus 1974
te Nijmegen

This dissertation is approved by:

Promotores: prof. dr. ir. A. Nijholt, University of Twente, The Netherlands
prof. dr. T. Dijkstra, Radboud University Nijmegen, The Netherlands

Assistent-promotor: dr. J. H. D. M. Westerink, Philips Research, The Netherlands

ISSN: 1381-3617; CTIT Ph.D.-thesis series No. 11-204
ISBN: 978-90-365-3243-3
DOI: 10.3990/1.9789036532433

Contents

List of Figures

List of Tables

I. Prologue

1

Introduction

Abstract

The quest towards an in-depth understanding of *affective computing* begins here. This is needed as advances in computing and electrical engineering seem to show that the unthinkable (e.g., huggable computers) is possible (in time). I will start with a brief general introduction in Section 1.1. Subsequently, Sections 1.2–1.4 will introduce three core elements of this monograph: i) Affect, emotion, and related constructs, ii) *affective computing*, and iii) *Affective Signal Processing (ASP)*. Next, in Section 1.5, the working model used in this monograph will be presented: a closed loop model. The model's signal processing and pattern recognition pipeline will be discussed, as this forms the (technical) foundation of this monograph. Section 1.6 will denote the relevance of *ASP* for computer science, as will be illustrated through three of its disciplines: human-computer interaction, artificial intelligence, and health informatics. This provides us with the ingredients for the quest for guidelines for *ASP* as described in this monograph. As such, I hope that this monograph will become a springboard for research on and applications of *affective computing*. I will end with an outline of this monograph.

Parts of this chapter are taken from:

Broek, E.L. van den, Nijholt, A., & Westerink, J.H.D.M. (2010). Unveiling Affective Signals. In E. Barakova, B. de Ruyter, and A.J. Spink (Eds.), *ACM Proceedings of Measuring Behavior 2010: Selected papers from the 7th international conference on methods and techniques in behavioral research*, Article No. a6. August 24–27, Eindhoven – The Netherlands.

and on the first three sections of:

Broek, E. L. van den, Janssen, J.H., Zwaag, M.D. van der, Westerink, J.H.D.M., & Healey, J.A. Affective Signal Processing (ASP): A user manual. *[in preparation]*

which already appeared partially as:

Broek, E.L. van den *et al.* (2009/2010/2011). Prerequisites for Affective Signal Processing (ASP) - Parts I-V. In A. Fred, J. Filipe, and H. Gamboa, *Proceedings of BioSTEC 2009/2010/2011: Proceedings of the International Joint Conference on Biomedical Engineering Systems and Technologies*. January, Porto, Portugal / Valencia, Spain / Rome, Italy.

1.1 Introduction

Originally, computers were invented for highly trained operators, to help them do massive numbers of calculations [149, 610]. However, this origin dates from the first half of the previous century, and much has changed since. nowadays, everybody uses them in one of their many guises. Whereas previously computers were stationary entities the size of a room, today we are in touch with various types of computers throughout our normal daily lives, including our smart phones [3, 122, 135, 381, 460, 610, 713]. Computation is on track to become even smaller and more pervasive. For example, microrobots can already flow through your blood vessels and identify and treat physical damage [2, 165, 214, 479]. Moreover, from dedicated specialized machines, computers have become our window to both the world and our social life [145, 472, 532].

Computers are slowly becoming dressed, huggable, and tangible and our reality will become augmented by virtual realities [50, 594]. Artificial entities are becoming personalized and are expected to understand more and more of their users' feelings, emotions, and moods [286, 594, 671]. Consequently, concepts such as emotions, that were originally the playing field of philosophers, sociologists, and psychologists [302] have become entangled in computer science as well [210]. This topic was baptized *affective computing* by Rosalind W. Picard [520, 521]. Picard identified biosignals as an important covert channel to capture human emotions, in addition to channels such as speech and computer vision.

Biosignals (or physiological signals) can be defined as (bio)electrical signals recorded on the body surface, although both non-electrical biosignals and invasive recording techniques exist as well. These bio(electrical) signals are related to ionic processes that arise as a result of electrochemical activity of cells of specialized tissue (e.g., the nervous system). This results in (changes in) electric currents produced by the sum of electrical potential differences across the tissue. This property is similar regardless of the part of the body the cells are located (e.g., the heart, muscles, or the brain) [245, 620]. For an overview of biosignals used for *affective computing*, I refer to Table 1.1.

There have been many studies that have investigated the use of biosignals for *affective computing* in the last decade. In Section 1.3 an overview of relevant handbooks will be provided and in Chapter 2 an exhaustive review of research articles will be provided. The handbooks and articles have in common that they illustrate, as I will also show later on (i.e., Chapter 2), that the results on *affective computing* have been slightly disappointing at best. Hence, I believe a careful examination of the current state-of-the-art can help to provide new insights for future progress. In sum, the goal of this monograph is to i) review the progress made on the processing of biosignals related to emotions (i.e., *Affective Signal Processing (ASP)*), ii) conduct necessary additional research, and iii) provide guidelines on issues that need to be tackled in order to improve *ASP*'s performance.

Table 1.1: An overview of common physiological signals and features used in ASP. The reported response times are the absolute *minimum*; in practice longer time windows are applied to increase the recording's reliability.

Physiological response	Features	Unit	Response time
Cardiovascular activity	Heart rate (HR)	beats / min	0.67-1.5 sec
through ElectroCardioGram (ECG)	SD IBIs, RMSSD IBIs	s	0.67-1.5 sec
or Blood Volume Pulse (BVP)	Low Frequency (LF) power (0.05Hz - 0.15Hz)	ms^2	0.67-1.5 sec
(per beat) [43, 44, 349]	High Frequency (HF) power (0.15HZ - 0.40Hz), RSA	ms^2	0.67-1.5 sec
	Very Low Frequency (VLF) power ($<$ 0.05Hz)	ms^2	0.67-1.5 sec
	LF/HF	ms^2	0.67-1.5 sec
	Pulse Transit Time (PTT)	ms	0.67-1.5 sec
Electrodermal Activity (EDA)	Mean, SD SCL	μS	after 2-10 sec
[62]	Nr of SCRs	nr / min	after 2-10 sec
	SCR amplitude	μS	after 2-10 sec
	SCR 1/2 recovery time, SCR rise time	s	after 2-10 sec
Skin temperature (ST)	Mean	^oC	after 15-20 sec
Respiration (per breath)	rate	nr / min	4-15 sec
[55, 238]	amplitude	a.u.	4-15 sec
	ins, exh	sec	4-15 sec
	total duty cycle	ins / cycle sec	4-15 sec
	ins exh	ins / exh sec	4-15 sec
Muscle activity [548]	Mean, SD EMG*	μV	$<$ 1 sec
through ElectroMyoGram (EMG)	Mean, SD inter-blink interval	ms	$<$ 1 sec
Movements / Posture [201, 403]	Alternating Current component (motion)	Hz	$<$ 1 sec
through Accelerometer [124, 190]	Direct Current component (posture)	Hz	$<$ 1 sec
Impedance Cardiography	Left-ventricular ejection time (LVET)	sec	per beat
[606, 623]	Prepreejection period (PEP)	sec	per beat
	Stroke Volume (SV)	ml	per beat
	Cardiac Output (CO)	liters/min	1 minute
	Total peripheral resistance (TPR)	MAP*80/CO	per beat
Blood Pressure (BP)	*both systolic and diastolic*	mmHg	per beat

Legend: SD: Standard deviation; RMSSD: Root Mean Sum of Square Differences; IBI: Inter-beat interval; ins: inspiration; exh: exhalation; RSA: Respiratory Sinus Arrhythmia; SCL: Skin Conductance Level; SCR: Skin Conductance Response.
* Most often the EMG of the corrugator supercilii, zygomaticus major, frontalis, and upper trapezius are used for ASP.

In the next section, I will provide a concise introduction on this monograph's core constructs affect, emotion, and related constructs. Subsequently, in Section 1.3, I will provide a concise overview of *affective computing* by providing both a definition of the field and a list of its representative handbooks. Section 1.4 will provide a definition of *Affective Signal Processing (ASP)* and will introduce its research rationale. Next, in Section 1.5, my working model for automatizing the recognition of emotion from biosignals will be introduced: a closed loop model for *affective computing*. One component of the model receives our main attention: the signal processing + pattern recognition processing pipeline. In Section 1.6, I will describe how the work presented in this monograph is embedded in computer science. Last, in Section 1.7, I will provide an outline of this monograph.

1.2 Affect, emotion, and related constructs

In 1993, Robert C. Solomon noted in the *Handbook of Emotions* [396, Chapter 1, p. 3, 1st ed.] that " *"What is an emotion?"* is the question that "was asked in precisely that form by William James, as the title of an essay he wrote for *Mind* well over 100 years ago (James, 1884). ...But the question *"What is an emotion?"* has proved to be as difficult to resolve as the emotions have been to master. Just when it seems that an adequate definition is in place, some new theory rears its unwelcome head and challenges our understanding." Regrettably, there is no reason to assume that this could not be the case for the concise theoretical framework that will be presented here (cf. [302]). Nevertheless, we need such a framework to bring emotion theory to *affective computing* practice.

In 2003, 10 years after Solomon's notion, in the journal *Psychological Review*, James A. Russell characterized the state-of-the-art of emotion (related) research as follows: *"Most major topics in psychology and every major problem faced by humanity involve emotion. Perhaps the same could be said of cognition. Yet, in the psychology of human beings, with passions as well as reasons, with feelings as well as thoughts, it is the emotional side that remains the more mysterious. Psychology and humanity can progress without considering emotion – about as fast as someone running on one leg."* [567, p. 145]. Where Solomon [396, Chapter 1, p. 3, 1st ed.] stressed the complexity of affect and emotions, Russell [567, p. 145] stressed the importance to take them into account. Indeed, affect and emotions are of importance psychology and humanity but *also* for (some branches of) science and engineering, as we will argue in this monograph.

Solomon's [396, Chapter 1, p. 3, 1st ed.] and Russell's [567, p. 145] quotes perfectly points towards the complexity of the constructs at hand (i.e., affect and emotion, amongst other things). It is well beyond the scope of this monograph to provide an exhaustive overview of theory on affect, emotion, and related constructs. However, a basic understanding and stipulative definitions are needed, as they are the target state *affective computing* and *ASP* are aiming at. This section will provide the required definitions. Since this monograph aims at *affective computing* and *ASP*, I will focus on affect as the key construct, which is, from a taxonomic perspective, a convenient choice as well. Affect is an umbrella construct that, instead of emotions, incorporates all processes I am interested in, as we will see in the remaining section.

Core affect is a neurophysiological state that is consciously accessible as a primitive, universal, simple (i.e., irreducible on the mental plane), nonreflective feeling evident in moods and emotions [531, 567]. It can exist with or without being labeled, interpreted, or attributed to any cause [567]. People are always and continuously in a state of core affect, although it is suggested that it disappears altogether from consciousness when it is neutral and stable [567]. Affect influences our attitudes, emotions, and moods and as such our feelings, cognitive functioning, behavior, and physiology [236, 567]; see also Table 1.2. As such, affect is an umbrella construct, a superordinate category [236].

Affect is similar to Thayer's activation [647], Watson and Tellegen's affect [707], and Morris' mood [462] as well as what is often denoted as a feeling [567]. As such, core affect is an integral blend of hedonic (pleasure-displeasure) and arousal (sleepy-activated) values; in other words, it can be conveniently mapped onto the valence-arousal model [372, 566, 567, 647]. However, note that the term "affect" is used throughout the literature in many different ways [531]. Often it is either ill defined or not defined at all. However, affect has also been positioned on another level than that just sketched; for example, as referring to behavioral aspects of emotion [236].

With affect being defined, we are left with a variety of related constructs. To achieve a concise but proper introduction to these constructs, we adopt Scherer's table of psychological constructs related to affective phenomena [58, Chapter 6]; see Table 1.2. It provides concise definitions, examples, and seven dimensions on which the constructs can be characterized. Together this provides more than rules of thumb, it demarcates the constructs up to a reasonable and workable level. Suitable usage of Table 1.2 and the theoretical frameworks it relies on opens affect's black box and makes it a gray box [323, 517], which should be conceived as a huge progress. The relations affective processes have with cognitive processes are also interested in this perspective. These will be discussed in Section 1.6.

1.3 Affective Computing: A concise overview

Affect and its related constructs (see Section 1.2) have already been a topic of research for centuries. In contrast, computers were developed only a few decades ago [149, 610]. At a first glance, these two topics seem to be worlds apart; however, as denoted in Section 1.1, emotions and computers have become entangled and, in time, will inevitably embrace each other. Their relation, however, is fresh and still needs to mature.

Table 1.2: Design feature delimitation of psychological constructs related to affective phenomena, including their brief definitions, and some examples. This table is adopted from [58, Chapter 6] and [219, Chapter 2].

construct	brief definition and *examples*	intensity	duration	synchronization	event focus	appraisal elicitation	rapidity of change	behavioral impact
Emotion	Relatively brief episode of synchronized response of all or most organismic subsystems in response to the evaluation of an external or internal event as being of major significance *(e.g., angry, sad, joyful, fewful, ashamed, proud, elated. desperate)*.	++ → +++	+	+++	+++	+++	+++	+++
Mood	Diffuse affect state, most pronounced as change in subjective feeling, of low intensity but relatively long duration, often without apparent cause *(e.g., cheerful, gloomy, irritable, listless, depressed, buoyant)*.	+ → ++	++	+	+	+	++	+
Interpersonal stances	Affective stance taken toward another person in a specific interaction, coloring the interpersonal exchange in that situation *(e.g., distant, cold, warm, supportive, contemptuous)*.	+ → ++	+ → ++	+	++	+	+++	++
Attitude	Relatively enduring, affectively colored beliefs, preferences, and predispositions towards objects or persons *(e.g., liking, loving, hating, valuing, desiring)*.	0 → ++	++ → +++	0	0	+	0 → +	+
Personality traits	Emotionally laden, stable personality dispositions and behavior tendencies, typical for a person *(e.g., nervous, anxious, reckless, morose, hostile, envious, jealous)*.	0 → +	+++	0	0	0	0	+

In 1995, Rosalind W. Picard wrote a technical report [520], which was a thought-paper that presented her initial thinking on *affective computing*. In a nutshell, this report identifies a number of crucial notions which are still relevant. Moreover, Picard provided an initial definition of *affective computing*: "...a set of ideas on what I call "affective computing," computing that relates to, arises from, or influences emotions." [520, p. 1]

In 2005, the first International Conference on Affective Computing and Intelligent Interaction (ACII) was organized. Two of the conference chairs, Tao and Tan, wrote a review on *affective computing* in which they defined it as: *"Affective computing is trying to assign computers the human-like capabilities of observation, interpretation and generation of affect features."* (cf. [639]). As such, they assured a one-on-one mapping of affect onto the traditional computer science / Human-Computer Interaction (HCI) triplet input (i.e., observation), processing (i.e., interpretation), and output (i.e., generation).

In 2010, the IEEE Transactions on Affective Computing were launched. Its inaugural issue contained a review by Rafael A. Calvo and Sidney D'Mello [87] who characterized the rationale of *affective computing* with: *"automatically recognizing and responding to a user's affective states during interactions with a computer can enhance the quality of the interaction, thereby making a computer interface more usable, enjoyable, and effective."*

For this monograph, however, we will define *affective computing* as: *"the scientific understanding and computation of the mechanisms underlying affect and their embodiment in machines"*. This definition is inspired by the short definition of Artificial Intelligence (AI) provided by the Association for the Advancement of Artificial Intelligence (AAAI)*. Drawing upon this definition, I have compiled an overview of books (see Table 1.3) that can be considered as handbooks on or related to *affective computing*. As such, Table 1.3 provides a representative overview of the work conducted in this field.

I have chosen to exclude M.Sc./Ph.D.-theses from Table 1.3. However, three Ph.D.-theses from the early years of *affective computing* should be mentioned: Jennifer A. Healey's (2000) *"Wearable and automotive systems for affect recognition from physiology"* [269], Maja Pantic's (2001) *"Facial expression analysis by computational intelligence techniques"* [509], and Marc Schröder's (2004) *"Speech and emotion research: An overview of research frameworks and a dimensional approach to emotional speech synthesis"* [588], which are complementary with respect to the signals used. Healey [269], Pantic [509], and Schröder [588] utilized respectively biosignals, computer vision technology, and the speech signal. In the next chapter, I will discuss this triplet in more depth. Additionally, the numerous (edited) volumes of Klaus R. Scherer and colleagues, starting with [581] and [583] up to the more recent [582] and [584], should be acknowledged. His work is of tremendous importance for *affective computing*; however, only a minority of his work includes a computing component [578].

*Association for the Advancement of Artificial Intelligence (AAAI)'s URL: `http://www.aaai.org/`

Table 1.3: An overview of 24 handbooks on affective computing. Selection criteria: i) on emotion and/or affect, ii) either a significant computing or engineering element or an application-oriented approach, and iii) proceedings, M.Sc.-theses, Ph.D.-theses, books on text-analyses, and books on solely theoretical logic-based approaches were excluded.

	author(s)	year	title
[521]	Picard	1997	Affective Computing
[153]	DeLancey	2002	Passionate engines: What emotions reveal about the mind and artificial intelligence
[656]	Trappl et al.	2003	Emotions in humans and artifacts
[193]	Fellous & Arbib	2005	Who needs emotions? The brain meets the robot
[455]	Minsky	2006	The Emotion Machine: Commonsense thinking, Artificial Intelligence, and the future of the human mind
[527]	Pivec	2006	Affective and emotional aspects of Human-Computer Interaction: Game-based and innovative learning approaches
[500]	Or	2008	Affective Computing: Focus on emotion expression, synthesis and recognition
[303]	Izdebski	2008	Emotions in the human voice, Volume I–III
[716]	Westerink et al.	2008	Probing Experience: From assessment of user emotions and behaviour to development of products
[558]	Robinson & el Kaliouby	2009	Computation of emotions in man and machines
[573]	Sander & Scherer	2009	The Oxford companion to emotion and affective sciences
[639]	Tao & Tan	2009	Affective Information Processing
[662]	Vallverdú & Casacuberta	2009	Handbook of research on synthetic emotions and sociable robotics: New applications in Affective Computing and Artificial Intelligence
[487]	Nishida et al.	2010	Modeling machine emotions for realizing intelligence foundations and applications
[526]	Pittermann et al.	2010	Handling emotions in human-computer dialogues
[533]	Prendinger & Ishizuka	2010	Life-like characters: Tools, affective functions, and applications
[582]	Scherer et al.	2010	Blueprint for Affective Computing: A sourcebook
[88]	Calvo & D'Mello	2011	New perspectives on affect and learning technologies
[228]	Gökçay & Yildirim	2011	Affective Computing and Interaction: Psychological, cognitive and neuroscientific perspectives
[218]	Fukuda	2011	Emotional engineering: Service development
[515]	Petta et al.	2011	Emotion-Oriented Systems: The Humaine handbook
[714]	Westerink et al.	2011	Sensing Emotions: The impact of context on experience measurements
[293]	Hudlicka	2012	Affective Computing: Theory, methods, and applications
[335]	Khosla et al.	2012	Context-aware emotion-based multi-agent systems

1.4 Affective Signal Processing *(ASP)*: A research rationale

As was already stated, this monograph focusses on *ASP* instead of *affective computing*. This gives rise to the question: what is the difference between the two? I have just provided a definition of *affective computing*. Hence, what is missing is a definition of *ASP*. In Section 1.1 of this chapter, *ASP* was briefly denoted as *"processing biosignals related to emotions"*.

This directly excludes the computer vision branch of *affective computing*, including vision-based analyses of facial expressions and body movements. Speech is not a direct biosignal either. However, it is an indirect biosignal, as will be explained in Table 2.2 of Chapter 2. This positions speech on the borderline of being a biosignal. However, the reasons just mentioned speak in favor of denoting speech as a biosignal. Therefore, in this monograph, for *ASP* purposes it is included as a biosignal.

In this monograph, the *signals* are: biosignals (or physiological signals) and speech. By *processing* these signals we mean signal processing + pattern recognition, as will be explained in Section 1.5. Processing these signals should result in the identification of people's *affective* states. Taken together, in this monograph, we adopt the following definition of *ASP: processing biosignals with the aim to acquire a scientific understanding and computation of the mechanisms underlying affect and their embodiment in machines.*

Now that *ASP* is defined, the question remains what the distinction is between *affective computing* and *ASP*? This is their difference in foci. In practice, research on *affective computing* often relies on its computing component (e.g., pattern recognition). With the adoption of *ASP* as research rationale instead of *affective computing*, I want to shift the focus from computing to a proper mapping of the underlying affective processes on the characteristics of the biosignals. The underlying assumption behind this shift in focus between *affective computing* and *ASP* is that the computing component of *affective computing* can only be successful if this mapping is well understood.

In the next section, I will define a closed loop model for *ASP* (but which also would suit *affective computing* nicely). This model will prove to be generic as *ASP* is envisioned to be applied in virtually all possible situations. Moreover, it allows us to discuss both *affective computing* and *ASP* in more depth than done so far.

1.5 The closed loop model

For over a century, closed loop models have been known in science and engineering, in particular in control theory [619] and electronics [477]. Closed loop models can be concisely defined as control systems with an active feedback loop. This loop allows the control unit to dynamically compensate for changes in the system. The output of the system is fed back

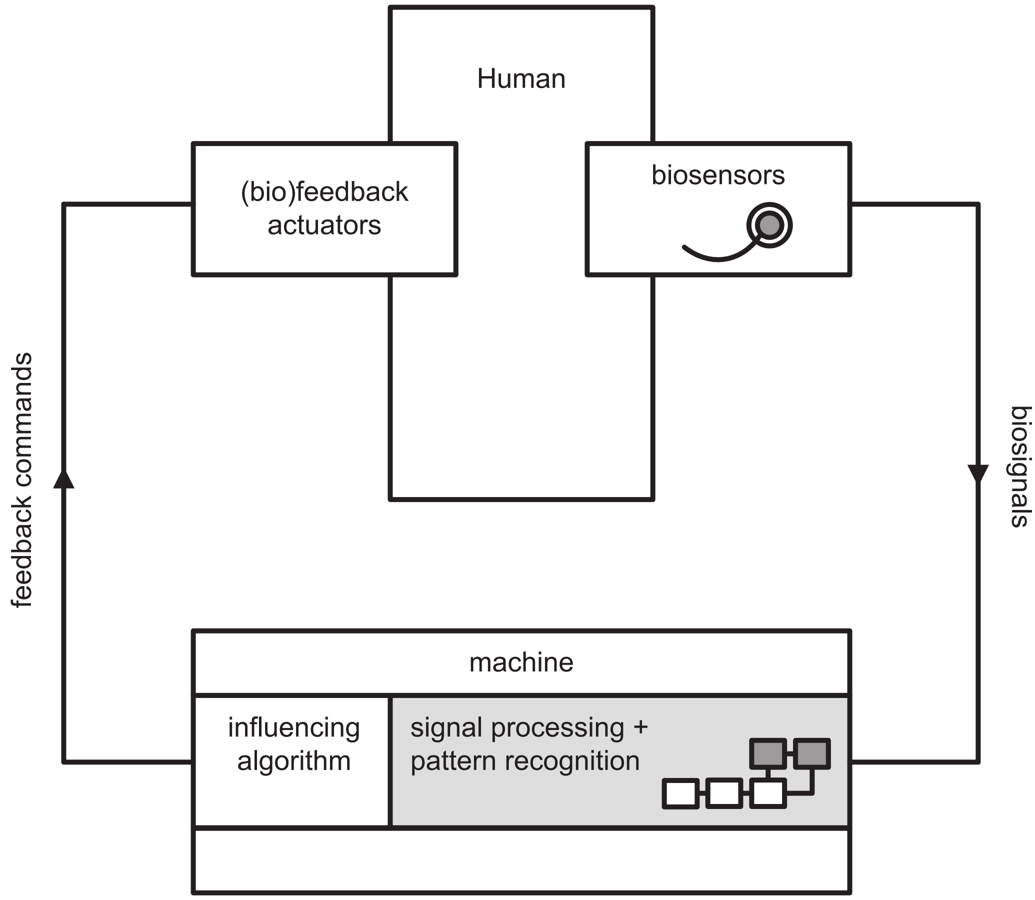


Figure 1.1: The (general) human-machine closed loop model. The model's signal processing + pattern recognition component, denoted in gray, is the component on which this monograph will focus (for more detail, see Figure 1.2). Within the scope of this monograph, the model's domain of application is *affective computing*.

through a sensor measurement to a control unit, which takes the error between a reference and the output to change the inputs to the system under control. In control theory, two types of control systems are distinguished: single-input-single-output (SISO) and Multi-Input-Multi-Output (MIMO; i.e., with more than one input/output) control systems.

More recently, a new class of closed loop models was initialized: closed loops that take a human / a user into the loop (cf. [587, p. 2]); see also Figure 1.1. Their descriptions target various areas but are essentially the same, comprising: sensors, processing, modeling, and actuators. We assume multiple inputs and outputs; hence, in terms of control theory, we introduce a new class of MIMO closed loop models. Their target state can be either one of the user or one of the system; that is, the user controlling the system or the system steering the user (in our case, to a certain emotional state). However, in the field of *ASP*, we assume the latter instead of the former. Recent application areas include Brain Computer Interfaces (BCI) [486, 637, 683], medical applications (e.g., sedation of patients [249] and rehabilitation [489]), and, as already mentioned, affective loops [83, 288, 654].

Since *affective computing* is approached from a range of sciences (e.g., psychology, medicine, and computer science), it is hard to provide a taxonomy for research on *affective computing*. However, it is feasible to identify the main types of research:

1. Computational modeling founded on theory, without experimental validation.
2. Emotion elicitation and measurement, with or without classification component. This type of research is conducted in three environments.
   (a) controlled laboratory research
   (b) semi-controlled research (e.g., as conducted in smart homes)
   (c) ambulatory research
3. Development of models, in which one can distinguish:
   (a) offline modeling
   (b) online, real-time modeling

This division is not as strict as it may appear, often mixtures of these types of research are employed. However, it should be noted that the vast majority of research on *affective computing* to date has not applied closed loop models, with McHugo and Lanzetta [83, Chapter 23] and, more recently, Tractinsky [654], Höök [288], and Janssen, Van den Broek, and Westerink [316] being exceptions. Instead most studies conducted either theoretical computational modeling or emotion elicitation and measurement. Moreover, most research has been conducted in (semi-)controlled settings. Ambulatory research with loose constraints, conducted in the real world, is still relatively rare (cq. [269, 270, 272, 316]). Nevertheless, I take the closed loop model as starting point and direct this monograph to ambulatory, real world *affective computing*.

Affective closed loops are important in *affective computing* applications that measure affective state and, subsequently, use these measurements to change the behavior of the systems. This allows computers to become more personal and social, and take into account how someone feels. Examples of affective closed loops are for instance a computer system that adapts its interaction dialogue to the level of frustration of its user [328, 576], or a music player that chooses the music to be played so as to guide the user to a better mood [316].

In essence, such affective closed loops are described by four basic steps (see Figure 1.1):

1. **Sensing:** Data collection starts at the sensors, where a raw signal is generated that contains an indication of a person's affective state. Relevant signals can include both overt and covert bodily signals, such as facial camera recordings, movements, speech samples, and biosignals (e.g., ElectroCardioGraphy (ECG) [100, 167, 317, 322, 375, 433, 434, 493, 494, 498, 513, 514, 585, 632, 738] or ElectroMyoGraphy (EMG) [133, 134, 206, 277, 446, 447, 664, 665, 667]).

2. **Signal processing + pattern recognition:** Exploiting signal features that could contain emotional information; for example, the number of peaks in the ElectroDermal Activity (EDA) [62, 136, 163, 203, 437, 497, 530, 536, 577] signal is counted, serving as a measure for arousal. Or the presence and strength of a smile can be derived from camera recordings, serving as measures for happiness.
3. **Influencing algorithm:** Given the obtained affective state of the user, a decision is made as to what is the best way to influence a user. These influencing algorithms need to incorporate a sense of what affective state the user wants or needs to reach (a goal state) as well as a model of how the user is likely to react to specific changes of the actuation system. Both serve to help the system in steering the user's emotional state.
4. **Feedback actuators:** The resulting emotion influencing is then undertaken by a set of actuators. Such actuators can directly communicate with our body, either physical [160, 265] or chemically [159, 451]. Alternatively, actuators can communicate indirectly and influence our environment as we sense it either consciously or unconsciously; for instance, a song can be played or lighting can be activated to create a certain ambiance.

The loop (always) closes when the sensors evaluate whether or not the intended emotional state has indeed been reached. If the intended emotional state indeed has been reached, the system will perform a NULL action.

Closed loop systems for *ASP* put a relatively large amount of emphasis on measurement, signal processing and pattern recognition. In general, two phases in this processing scheme are most often distinguished:

1. signal processing and
2. classification (e.g., in terms of emotion classes).

These two phases often form the core of the closed loop model, which can be considered as a signal processing + pattern processing pipeline, as is shown in Figure 1.2. Therefore, I will now first describe this general processing pipeline, before going back to the domain of *affective computing*.

Machines' learning of affect is essentially a signal processing + pattern recognition problem. The goal of pattern recognition techniques is to develop an artificial model that is able to recognize (complex) patterns, in our case emotions, through (statistical) learning techniques. It follows the classic pattern recognition processing pipeline (see also Figure 1.2 and [445]): a signal is captured and, subsequently, processed by a physical system (e.g., a CCD sensor, PC's audio card, or biosensor). After physical processing the raw signals provided (e.g., an image, audio track, or biosignal) form the *measurement space*.

The raw signals are preprocessed (e.g., filtered and artifacts removed), which provides 'clean' signals. After synchronization of these 'clean' signals, they can be segmented, based

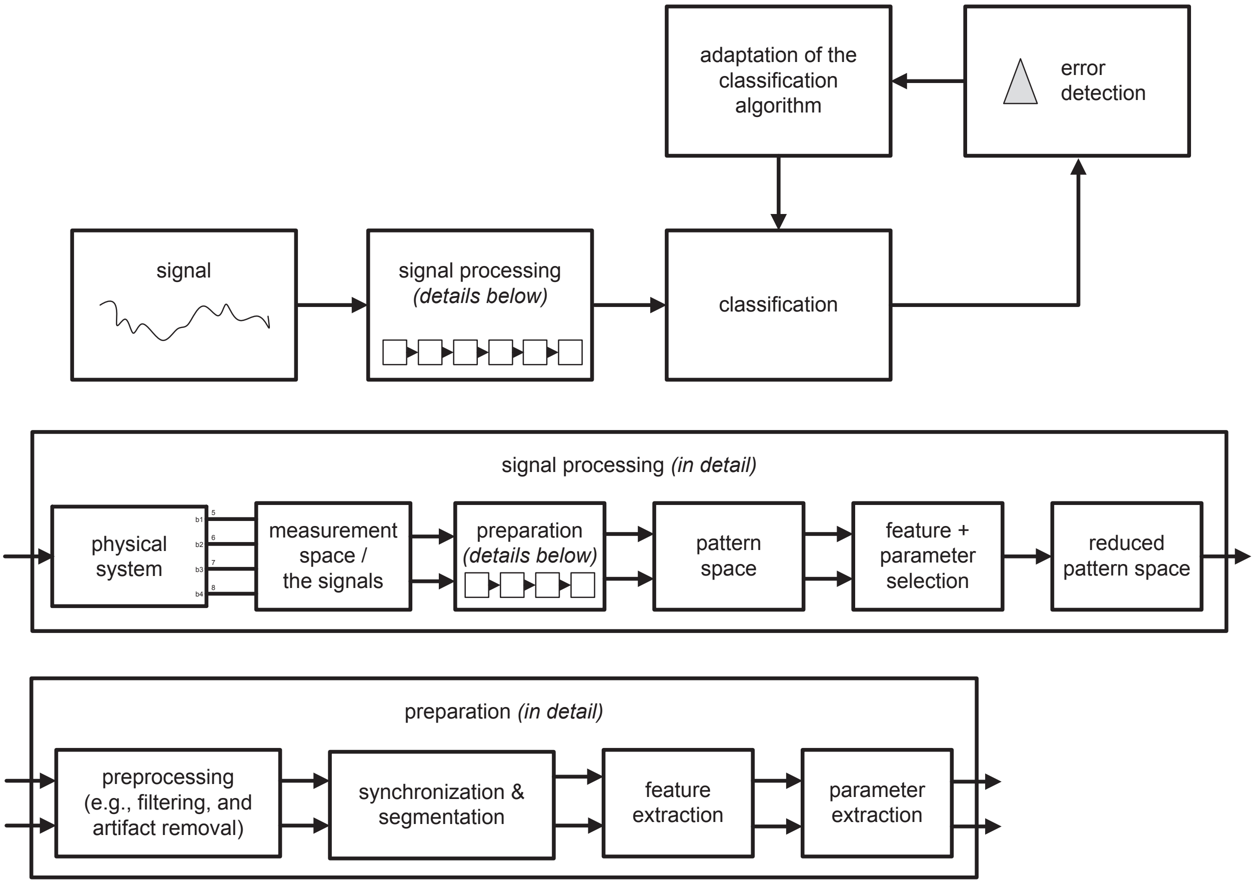


Figure 1.2: The signal processing + pattern recognition pipeline.

on events or stimuli, which facilitate their further analysis. Next, features need to be extracted from the signals and the parameters of these features need to be calculated. The affective signals are processed in the time (e.g., statistical moments [716, Chapter 14]), frequency (e.g., Fourier), time-frequency [51] (e.g., wavelets [143]), or power domain (e.g., periodogram and autoregression). In Table 1.1, I provide a brief overview of the signals most often applied, including their best known features, with reference to their physiological source. The set of features and their parameters provide the required *pattern space*.

The pattern space of calculated parameters from the recorded signals is defined for the pattern classification process. Next, feature selection / reduction is applied. This improves the prediction performance (or power) of the emotion classifier, reduces the chances of overfitting, provides faster and more cost-effective classifiers, and aids our understanding of the underlying process that generated the signals [243]. Consequently, the reduced parameter set eliminates the curse of dimensionality [48], removes redundancy between the signal's features and their parameters, and, hence, becomes more generic [68, 142, 708]. So, an optimal set feature vector (or more accurately: parameter vector) or reduced pattern space is generated, which can be fed to the classifier.

The next phase in the signal processing + pattern recognition is the actual classification of emotions using the optimized feature vectors. Three classes of pattern recognition techniques can be distinguished: statistical pattern recognition (including Artificial Neural Networks, ANNs [48, 266, 308]) (for more information, see also Appendix A), template matching (e.g., the work of Manfred Clynes [106–110, 480]), and syntactic or structural matching [75, 215, 216, 229, 655]. In *affective computing*, template matching, and syntactic or structural matching are seldom used, most often statistical classification is applied; see also Table 2.4. Statistical pattern recognition can be employed through either unsupervised or supervised classification (including reinforcement learning), which I will discuss next.

If a set of predefined classes (or labels or categories) to which the measurement space belongs is available (e.g., emotional states), the feature vector can be identified as a member of a predefined class and given the accompanying label. This approach is, therefore, baptized as supervised learning / classification (e.g., Fisher's Linear Discriminant Analysis, LDA and Support Vector Machines, SVM). Such predefined classes are sometimes referred to as the ground truth. In contrast, unsupervised classification techniques need to find structure in the data (e.g., Principal Component Analysis, PCA) or detect classes and class boundaries (e.g., clustering and LDA) without a ground truth (i.e., hitherto unknown classes) [226]. The classification process is based instead on the similarity of patterns, determined by a distance/similarity measure and an algorithm to generate the clusters of feature vectors representing an emotion.

In developing a classifying system, one can choose for either an unsupervised or a supervised approach [20]. Unsupervised classification does not need a priori knowledge and often only entails saving the pattern space in a specified format. Supervised classification requires the training (or learning) of a classifying system, before the actual classification can be conducted. Using labeled feature vectors for training, a discriminant function (or network function for ANN) is used to recognize the features and initial classification is realized. Classification errors can be determined using a certain error criterion and the classification process can be adapted [48, 170, 457, 648, 691]. This training or learning phase of supervised classification techniques is depicted by gray boxes in Figure 1.2, which are not applicable to unsupervised classification techniques.

This machine learning pipeline can be employed for each data source (i.e., modality such as vision, speech, and biosignals) separately. Alternatively, after the features and their parameters from all signals have been extracted, they can be merged into one pattern space. Both of these approaches are frequently applied. In the next chapter, I will discuss the pros and cons of each of the modalities and provide a review of each of them. Subsequently, an exhaustive review of biosignal-based *affective computing* will be provided. However, first I will end this chapter with sketching the relevance of *ASP* and *affective computing* for computer science and providing an outline of the rest of this monograph.

1.6 Three disciplines

The relation between emotions and computer science lays hold of various branches of computer science. The computer science disciplines that most noteworthy lay hold on emotions are:

1. HCI and related disciplines (e.g., user experience (UX) and interaction design, and cognitive ergonomics) [24, 264, 315, 545, 687, 735];
2. AI [292, 568], including agents and avatars [35, 38, 71, 217, 242, 297, 444, 686], robotics [69, 70, 672], and cognitive science and neuroscience [15, 147, 148, 162, 184, 431, 505]; and
3. Health Informatics, including e-health, and, more in general, health technology (including mobile technology) [18, 173, 296, 297, 321, 326, 702].

In the next three subsections, I will explain the relation of *ASP* with each of these three branches of computer science.

1.6.1 Human-Computer Interaction (HCI)

In the 90s of the previous century, Nass and colleagues [475, 549] touched upon a new level of HCI: a personal, intimate, and emotional level. Together with the work of Picard [520, 521] their work positioned affective processes firmly as an essential ingredient of HCI.

The importance of affect for HCI can be well explained by denoting its effects on three cognitive processes, which are important in HCI context:

1. **Attention**: Affective processes take hold on several aspects of our cognitive processing [118] and, hence, HCI [695]. One of the most prominent effects of affect lies in its ability to capture attention. Affective processes have a way of being completely absorbing. Functionally, they direct and focus our attention on those objects and situations that have been appraised as important to our needs and goals [695]. This attention-getting function can be used advantageously in HCI context [594, Chapter 4].
2. **Memory**: However, it should be noted that such effects also has implications for learning and memory [49, 64]. Events with an affective load are generally remembered better than events without such a load, with negative events being dominant over positive events [549]. Further, affect improves memory for core information, while undermining memory for background information [396, Chapter 37].
3. **Decision making**: Affective processes also have their influence on our flexibility and efficiency of thinking and problem solving [396, Chapter 34]. It has also been shown

that affect can (heavily) influence judgment and decision making [39, 593]. Affective processes tend to bias thoughts by way of an affect-filter. Thoughts are directed to a affect-consistent position, which can also increase the risk of distractions.

This triplet of cognitive processes illustrates that a careful consideration of affect in HCI can be instrumental in creating interfaces that are both efficient and effective as well as enjoyable and satisfying [594, Chapter 4].

1.6.2 Artificial Intelligence (AI)

Almost half a century ago, the American psychologist Ulric Neisser [478] stated that "*Human thinking begins in an intimate association with emotions and feelings which is never entirely lost*". Nobel prize winner and recipient of the ACM's Turing Award, Herbert A. Simon had similar ideas on this topic: "*...how motivational and emotional controls over cognition can be incorporated into an information-processing system ...'*" [611, p. 29].

Nonetheless, in the decades that followed AI aimed at understanding human cognition *without* taking emotion into account [568]. Although emotions were sometimes denoted as important (e.g., [454, 455]), it took until the publication of Picard's book *Affective computing* [521] before they received center stage attention. Even though AI has made it possible that a computer can beat the world's best chess players [89] and can win quizzes such as Jeopardy! [195], the general opinion is that AI has failed [426] (cf. [111, 332]). This is likely to be (partly) because of a lack of focus on emotions. So, almost 50 years after Ulric Neisser's words, with the user more demanding than ever, perhaps now is the time to bring emotions to the front line of AI research and practice.

1.6.3 Health Informatics

In 1935, Flanders Dunbar noted that the "*Scientific study of emotion and of the bodily changes that accompany diverse emotional experience marks a new era in medicine*". We now know that many physiological processes that are of profound significance for health can be influenced by way of emotions (e.g., [183, 305, 326]). For example, it has been shown that emotions influence our cardiovascular system [276, 400, 493, 495, 496, 589, 659] and, consequently, can shorten or prolong life [204, 205, 298, 299, 423, 598]. Moreover, emotions also play an important role with chronic diseases [42, 46], cancer (e.g., coping strategies) [343, 645], and rehabilitation [489], to mention three. [174, p. vii]. Nevertheless, emotions remained rather spiritual and human's health has usually been explained in physical (e.g., injuries) and physiological terms (e.g., bacteria and viruses). It is only since the last decades that it is generally acknowledged that emotions have their impact on health and illness [326, 493, 692].

Now they have been acknowledged by traditional medicine, emotions are now being given a position in health informatics. This shift was accelerated with the general increase in the need for health informatics that has emerged due to the massive growth of the market for new systems that improve productivity, cut costs, and support the transition of health care from hospital to the home [137, 173, 290, 296, 321, 393, 476, 605, 638, 702]. This transition relies heavily on (ethical) issues such as trust, persuasion, and communication [112] that have emotion as common denominator. Health informatics is already or will soon be applied for the support/assistance of independent living, chronic disease management, facilitation of social support, and to bring the doctor's office to people's homes. Par excellence, this is where health informatics and *affective computing* blend together.

1.6.4 Three disciplines, one family

The three disciplines described above are not mutually independent. For one thing, health informatics regularly applies AI techniques [568]; for example, the expert systems Eliza [711] and MYCIN [608] and their successors for various (sub)domains in medicine (e.g., [173, 677]). Also beyond medicine, expert systems have shown their added value, which is best illustrated by the fact that the user's role shifted from controller to supervisor [385]; for example, recently in high-end automobiles (e.g., Audi, BMW, and Mercedes-Benz). In all these cases, the systems interact with their users; hence, HCI takes a prominent place. This all stresses the relations between the three branches of computer science.

Besides the triplet discussed in this section, many other disciplines within computer science should take emotions into account as well; for example, virtual reality (VR) [86, 474, 488, 616], color processing [539], ambient intelligence (AmI) and ubiquitous computing (UbiComp) [1, 207, 471, 540, 668, 669, 676], multimedia [21, 595, 741], the World Wide Web (WWW) [199], and information retrieval (IR) [282, 283, 485]. To conclude, I hope that I have shown the substantial impact emotions have on many of the disciplines within computer science.

1.7 Outline

This monograph will be divided into five parts:

I A *prologue*,

II Basic research on *baseline-free ASP* that uses statistical moments,

III Basic research on *bi-modal ASP* that explores various aspects,

IV Three studies *affective computing*, and

V An *epilogue*.

A wide range of statistical techniques will be employed in the various chapters throughout this monograph. Appendix A will present these techniques in their simplest forms, will denote their characteristics, and will identify the relations between them.

I. Prologue: This part has started with the current chapter and will continue with:
Chapter 2, the second and last chapter of the prologue. In this chapter I will introduce *affective computing* and, more in particular, *ASP*. The three dominant modalities in this field (i.e., vision, speech, and biosignals) will be introduced. Next, I will provide the first exhaustive review on biosignal-based *affective computing*. Its advantages and disadvantages will be denoted as well as the reasons for the rapidly increasing interest in this modality.

II. Baseline-free *ASP*: This part will include two chapters in which I shall explore the feasibility of baseline-free *ASP* (i.e., without normalization in any phase of processing) using statistical moments:
Chapter 3. This chapter will cover research for which I used dynamic real world stimuli (i.e., movie scenes) to induce emotions. The EMG of three facial muscles was recorded, which is often done to establish a ground truth measurement. In addition, the participants' EDA was recorded. EDA is a robust and well documented biosignal that reveals the level of experienced arousal [62, 163].
Chapter 4. The research reported here consisted of analyses on the same data set as Chapter 3. The studies differed in the choice of time windows, which enabled research towards the impact and usage of this parameter for *ASP*. Where the analyses in Chapter 3 were executed on the complete signals accompanying the movie scenes, in this study 10 sec. time windows were used. Moreover, events in the movie scenes were annotated and the participants' affective responses that accompanied them were recorded.

III. Bi-modal *ASP*: Two studies will be presented that employed bi-modal *ASP* and deviate only with respect to the stimuli that were used for emotion elicitation. The research in these two chapters also assessed the influence of emotion representations by analyzing the obtained data using both the dimensional valence-arousal model [105, 176, 202, 452, 567, 647] and the six basic emotions [116, 181, 391]. Moreover, the impact of the environment (or context), the personality traits neuroticism and extroversion, and demographics on *ASP* was explored.
Chapter 5 will report research that employed a (or perhaps even *the*) reference set for *affective computing*: the International Affective Picture System (IAPS). The bi-modal *ASP* approach utilized the rare combination of ECG and speech. To the author's knowledge, this combination has further only been explored by Kim and colleagues [336, 337, 339, 340].
Chapter 6. In this chapter, I will present a study that is identical to the one in Chapter 5 except for the stimuli that have been applied to induce emotions in the participants. The type and selection of stimuli has recently (again) been shown to be a factor of importance [8]. In this study the same set of movie fragments was used as in Chapters 3 and 4. This enabled a

comparison of static versus dynamic stimuli and, as such, assessed their validity.

IV. Towards affective computing: In these three chapters I will explore the feasibility of *affective computing* using *ASP*.
Chapter 7. In this chapter, I will go through the complete signal processing + pattern recognition pipeline (see Figure 1.2), using the data that is also presented in Chapters 3 and 4. In the quest for an optimal processing pipeline, several signal processing aspects and classification methods (see also Appendix A) will be explored. As such, the feasibility of emotion-aware systems will be assessed.
Chapter 8. In this chapter two studies will be presented that bring us from lab research to clinical practice. For these studies, I employed only the speech signal since direct biosignals were considered to be too obtrusive. The studies' aim was to lay a foundation for the development of a Computer-Aided Diagnosis (CAD) of patients with a Post-Traumatic Stress Disorder (PTSD).
Chapter 9. In this chapter the data from the two studies presented in Chapter 8 will be fed to the same complete signal processing + pattern recognition pipeline as was already employed in Chapter 7. This explores the true feasibility of the envisioned emotion-aware systems, in this case: *ASP*-based Computer-Aided Diagnosis (CAD) for mental health care.

V. Epilogue: This part consists of a set of guidelines for *ASP* and a general discussion.
Chapter 10. In this chapter I will present the lessons learned while working on the research presented in this monograph. Considerations and guidelines for processing affective signals and classifying the features derived from these signals in terms of emotions will be introduced. These guidelines will indicate possible problems, will provide solutions for them, and will provide research directives for *affective computing*. As such, this will perhaps be the most important chapter of this monograph.
Chapter 11 will consist of seven sections. First, I will look back on the work conducted and draw some brief conclusions from this. Second, I will place the work presented in this monograph in a historical perspective. Third, I will weight this monograph's contribution to emotion science's 10 hot topics as has been recently identified [236]. Fourth, I will introduce *affective computing's I/O*. Fifth, I will describe three consumer applications that can be developed here and now! Sixth, I will stretch the horizon and describe two visions of the future: robot nannies and digital human models. Seventh and last, I will draw some final conclusions and close the monograph.

2

A review of Affective Computing

Abstract

Research on *affective computing* is dominated by three signal types: vision, speech, and biosignals. As will be shown through a concise review, the former two have a rather established background. The latter signal is less well explored and will, therefore, be exhaustively reviewed. In total, the chapter will review 85 articles, which have been published on this topic throughout more than a decade of research. From this structured review one should conclude that the field suffers from a lack of progress. The main reason for this seems to be the lack of standards in combination with the field's intrinsic complexity. This notion will serve as the foundation for this monograph's work, as will be shown in the next chapters.

This chapter is based on the fourth section of:

Broek, E. L. van den, Janssen, J.H., Zwaag, M.D. van der, Westerink, J.H.D.M., & Healey, J.A. Affective Signal Processing (ASP): A user manual. *[in preparation]*

which has already partially appeared as:

Broek, E.L. van den *et al.* (2009/2010/2011). Prerequisites for Affective Signal Processing (ASP) - Parts I-V. In A. Fred, J. Filipe, and H. Gamboa, *Proceedings of BioSTEC 2009/2010/2011: Proceedings of the International Joint Conference on Biomedical Engineering Systems and Technologies*. January, Porto, Portugal / Valencia, Spain / Rome, Italy.

2.1 Introduction

The closed loop model for *affective computing* (see Figure 1.1) requires an integrated signal processing and pattern recognition processing pipeline, as was outlined in Section 1.5 of the previous chapter (see also Figure 1.2). Where the last section in the previous chapter provided a general outline, this chapter will start with a brief discussion of the three *ASP* modalities (i.e., vision, speech, and biosignals) in more detail. In parallel, for both vision- and speech-based *affective computing* a concise review will be presented. Subsequently, an exhaustive review of biosignal-based *affective computing* will be presented (see Table 2.4), which is the focus of the current monograph. Results in this subfield will be discussed and conclusions will be drawn. These provide the ingredients for this monograph's stipulative definition of *ASP*.

2.2 Vision

Vision-based emotion recognition is popular. This has several reasons, among which the following triplet: i) vision is an intuitively attractive modality, which uses a common, inexpensive, non-contact sensor, ii) Ekman and Friesen's Facial Action Coding System (FACS) [180] provides an excellent set of markers that can be employed for vision-based emotion recognition [574], and iii) vision-based emotion recognition is feasible in controlled conditions. Vision-based emotion recogntion can be employed both as a static (i.e., image, either 2D or 3D [574]) and a dynamic technique (i.e., video) [241, 717, 727, 739]. Moreover, one can choose to conduct markerless recordings or to use markers, which facilitate processing the images or videos. Most research on *affective computing* incorporates recordings of the face but body language analysis has also recently been shown to be a rich source of information (e.g., movements and gestures) [28, 241, 351, 421, 720].

In the rich body of the literature, it has been shown that thousands of features can be derived from human faces [727]. These features can be characterized through three dimensions [192]: local features versus the face as a whole, deformation analysis versus motion extraction, and image versus model based approaches. In practice, the number of features extracted varies considerably: from 16 [732] to 4320 [727]. In most studies, feature selection/reduction is applied. The number of subjects that participated in the studies also varies considerably, from 4 to 210. In contrast, the number of emotion classes between which is discriminated is rather constant throughout the studies (range: $5 - 8$) with only one study that discriminates between 12 emotional states [241]. In line with this, the reported recognition rates are fairly constant over studies, ranging from 72% to 98%. The early work of Cottrell and Metcalfe [127] on *EMPATH* is an exception to this, with their 20% correct classification between 8 emotions (chance level: 12.5%).

Table 2.1: Review of 12 representative machine learning studies employing computer vision to recognize emotions.

source	year	signals	# subjects	features	selection / reduction	classifiers	target	classification
[127]	1991	face	20/5	64×64		ANN	8 emotions	20%
[185] [186]	1995 1997	FACS	8	physical model		CTKF	5 emotions	98%
[732]	1996	mouth motions	32	16			7 emotions	65%
[399]	2000	FACS	100	38		PCA,LDA,HMM	3-9 action units	>80%
[115]	2003	motion units	5/53	12		NB	7 emotions	73%/83%
[743]	2005	FACS		24		BN	6 emotions	72%
[510]	2006	FACS	19	24 rules		Particle filtering	30 action units	87%
[408]	2006	23 expressions	100	900	AdaBoost	LDA,SVM	7 emotions	93%
[240]	2007	head	4	148	BFS	BN	6 emotions	75%
		body		140				90%
		head & body		288				80%-94%
[241]	2009	head,hands,& schoulders	10	150/172	PCA,HMM,EM	DT,BN,SVM, ANN,AdaBoost	12 emotions	83%/85%
[572]	2010	face	41/52	84	GP,grid	SVM	6 emotions	92%/95%
[727]	2011	face	210	4320	GA,LPP	ANN,SVM	6 emotions	65%-97%

Legend: FACS: Facial Action Coding System [180]; AdaBoost: Adaptive Boosting; ANN: Artificial Neural Network; BN: Bayesian Network; BFS: Best First Search; CTFK: Continuous Time Kalman Filter; DT: Decision Tree; EM: ExpectationŰ-Maximization algorithm; GA: Genetic Algorithm; GP: Gaussian Process classification; HMM: Hidden Markov Models; LDA: Fisher's Linear Discriminant Analysis; LPP: Locality Preserving Projection; NB: Naïve-Bayes; PCA: Principal Component Analysis; SVM: Support Vector Machine.

Although vision-based facial expression analysis provides (reasonably) good results nowadays, it is complex. Physiognomies of faces vary considerably between individuals due to age, ethnicity, gender, facial hair, cosmetic products, and occluding objects (e.g., glasses and hair). Furthermore, a face can appear to be distinct from itself due to pose and/or lighting changes. For a more elaborate discussion on these issues, I refer to the surveys of Fasel and Luettin [192] and Tian, Kanade, and Cohn [652]. For a comparative evaluation of 2D versus 3D FACS-based emotion recognition, I refer to the recent study of Savran, Sankur, and Bilge [574]. In sum, in highly controlled environments, vision-based emotion recognition works properly; however, when constraints are loosened, it might fail [192, 652, 680, 717]. This makes it a powerful but also fragile affective signal.

2.3 Speech

Although speech-based emotion recognition shows many parallels with vision-based emotion recognition, it also has some distinct characteristics; see also Tables 2.2 and 2.3. A prominent difference is that speech-based emotion recognition has been applied on both healthy subjects *and* on subjects with varying disorders (e.g., [677]). Research on speech-based emotion recognition with people with disorders has revealed a lot of information on emotional speech.

For decades audio-based emotion recognition has been conducted with a limited set of features (≤ 64; see also Table 2.3), without the use of any feature selection or reduction [579, 696]. During the last decade, more often a brute force strategy was employed [590], using hundreds or even thousands of features (e.g., see [644, 725]). In parallel with the explosion of the number of features, feature selection/reduction strategies claimed their position. Machine's recognition rate of emotional speech ranges from Banse and Scherer [27], who report $25\%/40\%$ correct classification of 14 emotions, to Wu, Falk, and Chan [725], who report $87\%/92\%$ correct classification of 7 emotions. The latter results, however, are in contrast with the results on a structured benchmark reported by Schuller, Batliner, Steidl, and Seppi [590] at the InterSpeech 2009 emotion challenge: $66\% - 71\%$ (2 classes) and $38\% - 44\%$ (5 classes).

Similarly to vision-based approaches, audio-based emotion recognition suffers from environmental noise (e.g., from a radio or conversations in the background). Moreover, audio recordings are influenced by acoustic characteristics of environments; using templates for distinct environments could relieve this burden. Moreover, recording of sound could also cancel out some of the noise, although such an approach is in itself already very challenging. Similarly to vision-based emotion recognition, audio-based emotion recognition can be best employed in a controlled setting, as is the case in most studies.

Table 2.2: Speech signal analysis: A sample from history.

Throughout the previous century, extensive investigations have been conducted on the functional anatomy of the muscles of the larynx [281, 386]. It was shown that when phonation starts, an increase in electrical activity emerges in the laryngeal muscles. Also with respiration, slight electrical activity was found in the laryngeal muscles. These processes are highly complex as speech is an act of large motor complexity, requiring the activity of over 100 muscles [386]. These studies helped to understand the mechanisms of the larynx during phonation; cf. [653]. Moreover, algorithms were developed to extract features (and their parameters) from the human voice. This aided further research towards the mapping of physical features, such as frequency, power, and time, on their psychological counterparts, pitch, loudness, and duration [114].

In the current research, the physical features are assessed for one specific cause: stress detection. One of the promising features for voice-induced stress detection is the fundamental frequency (F0), which is a core feature in this study. The F0 of speech is defined as the number of openings and closings of the vocal folds within a certain time window, which occurs in a cyclic manner. These cycles are systematically reflected in the electrical impedance of the muscles of the larynx. In particular, the cricothyroid muscle has been shown to have a direct relation with all major F0 features [117]. In addition, it should be noted that F0 has a relation with another, very important, muscle: the heart. It was shown that the F0 of a sustained vowel is modulated over a time period equal to that of the speaker's heart cycle, illustrating its ability to express one's emotional state [501].

Through recording of speech signals, their features (e.g., amplitude and F0) can be conveniently determined. This has the advantage that no obtrusive measurement is necessary. Only a microphone, an amplifier, and a recording device are needed. Subsequently, for the determination of F0, appropriate filters (either hardware or software) can increase the relative amplitude of the lowest frequencies and reduce the high and mid-frequency energy in the signal. The resulting signal contains little energy above the first harmonic. In practice, the energy above the first harmonic is filtered, in a last phase of processing.

Harris and Weiss [263] were the first to apply Fourier analysis to compute the F0 from the speech signal. Some alternatives for this approach have been presented in the literature; for example, wavelets [712]. However, the use of Fourier analysis has become the dominant approach. Consequently, various modifications on the original work of Harris and Weiss [263] have been applied and various software and hardware pitch extractors were introduced throughout the second half of the 20^{th} century (cf. [168] and [537]). For the current study, we adopted the approach of Boersma [53] to determine the F0 of speech.

Table 2.3: Review of 12 representative machine learning studies employing speech to recognize emotions.

source	year	# subjects	features	selection / reduction	classifiers	target	classification
[398]	1962	3			10 humans	8 utterances	85%
[369]	1985	65	12/8/4	3 experiments		7 scales	n.a.
[27]	1996	12			12 humans	14/11 emotions	48%/55%
			18		LRM,LDA	14 emotions	25%/40%
[491]	2003	12			3 humans	6 emotions	66%
			64	distortion	HMM		78%
[503]	2003	6			8 humans	5 emotions	57%/77%
			200/ 20/20	GA,NB, k-NN	k-NN, DT, ANN, K*, LRM, SVM, VFI, DT, NB, AdaBoost	4 emotions	50%-96%
[463]	2007	11/12	38	PCA,SFS,GA	SVM,k-NN,ANN,NB,K*,DT	2/6 emotions	71-79%/60-73%
[580]	2009	10	>200	k-means	k-NN	7 emotions	70%
					20 humans		85%
[424]	2010	10	383	J_1 criterium,LDA	SVM,GMM	7 emotions	$\leq$78%
[644]	2010	10	1054	SFS	SVM	7 emotions	83%/85%
		4				3 emotions	88%/92%
[618]	2010	10	173	various	DT,SVM	8 emotions	75%/83%
[677] and Chapter 8	2011	25	65	SBS	LRM	SUD scale	83%
							69%
[725]	2011	10	442	SFS,LDA	SVM	7 emotions	87%-92%
		19		SFS	correlation	9 emotions	$\leq$81%
					human		$\leq$65%
		28		SFS	correlation		$\leq$69%
					human		$\leq$56%
		19+28		SFS	correlation		$\leq$73%
					human		$\leq$61%

Legend: AdaBoost: Adaptive Boosting; ANN: Artificial Neural Network; DT: Decision Tree; GA: Genetic Algorithm; GMM: Gaussian Mixture Models; HMM: Hidden Markov Models; K*: Instance-based classifier, with a distance measure based on entropy; k-NN: k-Nearest Neighbors; LDA: Fisher's Linear Discriminant Analysis; LRM: Linear Regression Model; NB: Naïve-Bayes; PCA: Principal Component Analysis; SBS: Sequential Backward Selection; SFS: Sequential Forward Selection; SVM: Support Vector Machine; VFI: Voted Features Interval.

2.4 Biosignals

Physiological signals and their relation to cognition and emotion have been topics of research for over two centuries [311, 313, 366]. In particular in the previous century interest from science and, subsequently, industry rose on the relation between physiological signals and emotion. A broad range of physiological signals are used for emotion recognition, see also Table 1.1. The choice of the signal(s) depends heavily on both the area of application (e.g., can it be integrated into another device or not) and on the information that needs to be extracted from it. In practice, most physiological signals are derived through non-invasive methods and, as such, are indirect measures (cf. [212, 213]). Therefore, often a delay between the actual physiological origin and the recorded change in the physiological signal has to be taken into account. Moreover, physiological signals differ significantly between individuals. Consequently, personalized approaches have been shown to have the best performance (cf. [338]).

Biosensors can also be unreliable due to movement artifacts and differences in bodily position [269, 270, 272, 466]. Even more problematic is that people's physiology is influenced by internal sources (e.g., a thought), a broad range of possible external factors (e.g., a signal outside [198]), and physical activity [78, 196]. The latter is illustrated in Figure 2.1. This makes affective signals inherently noisy (cf. Figure 2.1), which is most prominent in real world, ambulatory research [151, 269, 270, 272, 466, 632, 674]. To deal with this these factors should be measured and modeled as far as is possible (cf. [383]). Another issue is that the sensors have to be directly attached to human skin and, consequently, can be experienced as obtrusive (e.g., as is even still the case with state-of-the-art facial EMG [378, 664]), which can complicate their integration into real-world consumer applications. This has been the traditional burden for the application of *ASP* in end consumer products. However, with the rapid development of wireless networks, miniaturized sensors, body area networks, and reconfigurable biosignal recording devices (e.g., [3, 122, 135, 381, 460, 610, 713]), this traditional disadvantage is quickly vanishing [98, 121, 321, 403, 508, 571, 658, 702] (cf. [642]). Physiological signal recording apparatus has already been integrated in tools such as helmets, beds, music players, gaming consoles, or clothes [10, 101]. Consequently, physiological recordings are gaining in popularity and should be considered as promising.

2.4.1 A review

Few comprehensive reviews have appeared on *affective computing* using physiological signals, when compared with the other modalities. This also illustrates that research on *affective computing* using this modality is rather new. It started with the publication of Picard's book *Affective computing* in 1997. At that moment, audio-based emotion recognition, in particular speech-based emotion recognition had already been employed for decades. More

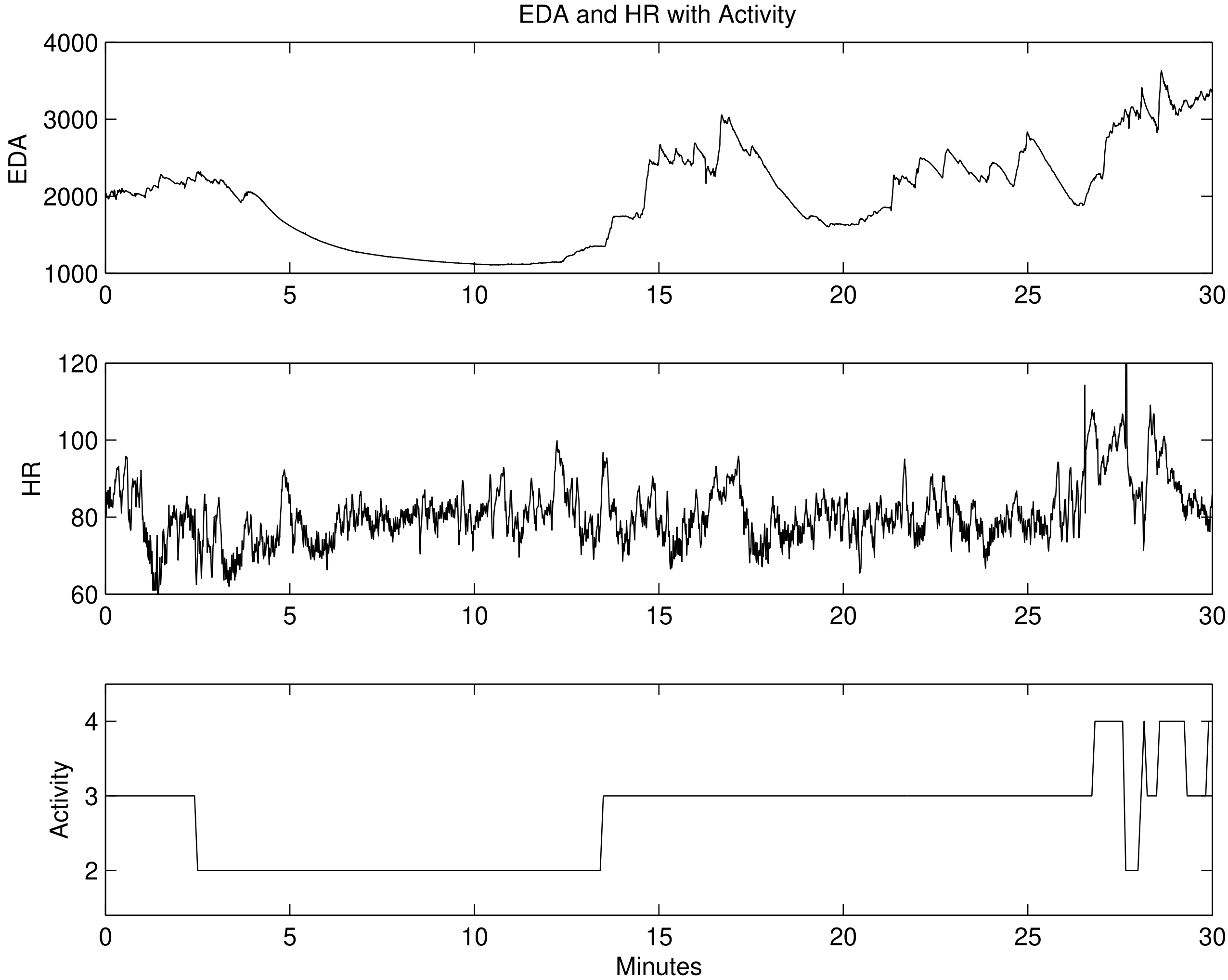


Figure 2.1: Recordings of Heart Rate (HR), ElectroDermal Activity (EDA), and a person's activity for a period of 30 minutes, in a real world setting.

recent, but still 10 years before the publication of Picard's book, vision-based emotion recognition was being employed. Even early works on multimodal emotion recognition had already been published, such as Tartter [640], although it should be mentioned that his classifiers were humans. Recently, a concise review appeared [352], which briefly wraps up some key notions of *affective computing*. They report 92% correct classification rate as best result, using (only) 4 signals and discriminating between 4 emotions. They pose that "*A recognition accuracy of over 80% on the average seems to be acceptable for realistic applications*" (p. 153). This claim is a rather bold statement, as in other (traditional) application areas of pattern recognition it has been shown that this is not the case. With most other pattern recognition problems, recognition rates of over 90% (and often over 95%) are achieved [308]; for example, multimedia analysis [324], optical character recognition (OCR) [461], and handwriting recognition [528], writer identification [73], and face recognition [65]. This illustrates the complex nature of *affective computing* as well as the need for standardization and benchmarks in *affective computing*.

Table 2.4 presents an extensive review of the research conducted in the biosignal branch of *affective computing*. One of the earliest works that should not remain unmentioned is that of Sinha and Parsons [613] who applied facial EMG on 27 subjects, extracted 18 features from this and achieved an 85% correct classification rate of 2 emotions, using a linear discriminant analysis (LDA). In 2001, Picard, Vyzas, and Healey [524] published their pioneering study with 81% correct classification of 8 emotions, also using LDA. Their study included multiple physiological signals but only one subject to which the complete signal processing and pattern recognition pipeline (see Figure 1.2) was tailored. In the decade that followed on this study, both the classification rate of Picard et al. [524] and the number of emotions between which they discriminated was not successfully challenged. An exception to this is the work of Kim and André [338] who reported 70% correct classification of 4 emotions using a generic classifier and 95% correct classification when employing a personalized classifier (see also [340]). This result supports the idea that one of the reasons the recognition rate by Picard et al. [524] could not be improved was that it used a single person classifier instead of a generic model. Most studies discriminated between $2-4$ emotion (categories) and achieved a correct classification in the range of 60% to over 90%; see Table 2.4.

More than anything else, Table 2.4 illustrates the variety in research on *affective computing*. Both the type and number of signals employed varies considerably. Distinct studies are hard to compare because they are executed in different settings, ranging from controlled lab studies to real-world, ambulatory testing. Also, the number of people participating varies from 1 to 50, although studies including > 20 participants are relatively rare. Moreover, features and parameters determined through *ASP* vary from 3 to 225. Last, only half of the studies applied feature selection/reduction, where this would generally be advisable. Moreover, a broad plethora of classifiers is used. The characteristics of the categories among which has to be discriminated is different from most other classification problems. The emotion classes used are typically ill defined, which makes it hard to compare studies. Moreover, the type and number of emotion categories (i.e., the classes) to be discriminated varies considerably (from 2 to 8) as well as the methods of elicitation. Consequently, although these are small numbers in terms of pattern recognition and machine learning, the results are not as good as those of other classification problems. With *affective computing* recognition rates of $60\% - 80\%$ are common, where in most other pattern recognition problems, recognition rates of $>> 80\%$ and often $> 95\%$ are frequently reported. This illustrates the complex nature of *affective computing* and the need to consider prerequisites for *ASP*.

Table 2.4: An overview of 61 studies on automatic classification of emotions, using biosignals / physiological signals.

source	year	signals	# subjects	features	selection / reduction	classifiers	target	classification
[613]	1996	$\mathcal{M}$	27	18		LDA	2 emotions	86%
[194]	1997	$\mathcal{C}$,$\mathcal{E}$	24	5	B-W	HMM,Viterbi	frustration / not	63%
[271]	1998	$\mathcal{C}$,$\mathcal{E}$,$\mathcal{R}$,$\mathcal{M}$	1	11	LDA	QuadC,LinC	3 emotions	87%-75%
							anger / peaceful	99%
							2 arousal levels	84%
							2 valence levels	66%
[703]	1998	$\mathcal{C}$,$\mathcal{E}$,$\mathcal{M}$,$\mathcal{R}$	1	24	SFFS,LDA	k-NN	8 emotions	$\leq$51%
							7 emotions	$\leq$54%
							5 emotions	$\leq$71%
							4 emotions	$\leq$73%
							3 emotions	$\leq$83%
[268]	2000	$\mathcal{C}$,$\mathcal{E}$,$\mathcal{R}$,$\mathcal{M}$	1	12	SFFS	k-NN	4 stress levels	87%
[524]	2001	$\mathcal{C}$,$\mathcal{E}$,$\mathcal{R}$,$\mathcal{M}$	1	40	SFFS	LDA	8 emotions	81%
[405]	2002	$\mathcal{C}$,$\mathcal{E}$,$\mathcal{T}$	10	3		k-NN,LDA	5 emotions	85%**
[576]	2002	$\mathcal{C}$,$\mathcal{E}$	24	5	Viterbi	HMM	2 frustrations	64%
[341]	2002	$\mathcal{C}$,$\mathcal{E}$,$\mathcal{T}$	50	10	-	SVM	4 emotions	61%
			(125)	10	-	SVM	3 emotions	55%/78%
[473]	2003	$\mathcal{C}$,$\mathcal{E}$,$\mathcal{T}$	31	3		k-NN,LDA	6 emotions	69%
[636]	$\mathcal{C}$,$\mathcal{E}$,$\mathcal{B}$	12	18		SVM	6 emotions	42%	
[636]	2003	$\mathcal{C}$,$\mathcal{B}$	10	12		NN,SVM	2 valence levels	62%
[544]	2003	$\mathcal{C}$,$\mathcal{E}$,$\mathcal{M}$,$\mathcal{T}$	1	18		FL, RT	3 anxiety levels	59%-91%
[246]	2004	$\mathcal{C}$,$\mathcal{E}$,$\mathcal{R}$,$\mathcal{M}$,$\mathcal{T}$	1	13		NN	arousal	90%–97%*
							valence	64%–90%*

source	year	signals	# subjects	features selection / reduction		classifiers	target	classification
[406]	2004	$\mathcal{C},\mathcal{E},\mathcal{T}$	29	12		k-NN,LDA,NN	6 emotions	84%
[342]	2004	$\mathcal{C},\mathcal{E},\mathcal{T}$	50	10	LDA	SVM	4 emotions	62%
			175	10		SVM	3 emotions	78%
[543]	2004	$\mathcal{C},\mathcal{E},\mathcal{M}$	1	6		FL	2 anxiety levels	??%
[275]	2004	$\mathcal{C},\mathcal{E},\mathcal{R},\mathcal{M},\mathcal{T}$	1	30	reg.,LDA	k-NN	5 emotions	24%
[387]	2004	$\mathcal{C},\mathcal{E},\mathcal{R},\mathcal{M}$	1	15	DBI	NN	2 classes	100%
[389]	2004	$\mathcal{C},\mathcal{E}$	8		DBI	NN	pos.&neg.	??%
[634]	2004	$\mathcal{C},\mathcal{E},\mathcal{B}$	12	18		SVM	5 emotions	42%
[635]			12	18		SVM	3 emotions	67%
[704]	2005	$\mathcal{C},\mathcal{E},\mathcal{R},\mathcal{M}$	1	32	SFS,LDA	k-NN,LDA,NN	4 emotions	92%
[736]	2005	$\mathcal{C},\mathcal{E}$	6	5		NN	4 emotions	80%
[104]	2005	$\mathcal{E}$	1	3	PCA	NN	4 emotions	75%
[272]	2005	$\mathcal{C},\mathcal{E},\mathcal{R},\mathcal{M}$	9	22	LDA	LDA	3 stress levels	97%
[339]	2005	$\mathcal{C},\mathcal{E},\mathcal{R},\mathcal{M}$	3	26	SFS	LDA	4 emotions	53%/74%
[407]	2005	$\mathcal{C},\mathcal{E},\mathcal{T}$	41	86		k-NN,NN (2x)	3×2 emotions	92%
[411]	2005	$\mathcal{C},\mathcal{E},\mathcal{M}$	15	13(?)		k-NN,RT,BN,SVM	5 emotions	86%
[745]	2005	$\mathcal{C},\mathcal{E},\mathcal{T}$,other	32	?		k-NN,NN	2 fear levels	92%
[538]	2006	$\mathcal{C},\mathcal{R}$	15	18	ANOVA,PCA,SFS	LDA	4 emotions	65%
							2 emotions	72%-83%
[412]	2006	$\mathcal{C},\mathcal{E},\mathcal{M},\mathcal{T}$	14	35		RT	3 anxiety levels	70%
[541]	2006	$\mathcal{C},\mathcal{E},\mathcal{M},\mathcal{T},\mathcal{P}$	15	46		k-NN,SVM,RT,BN	3 emotions	86%
[740]	2006	$\mathcal{C},\mathcal{E},\mathcal{T},\mathcal{P}$	32	11		SVM	2 stress levels	90%
[388]	2007	$\mathcal{C},\mathcal{E}$	8	5	DBI	NN	3 emotions	71%
[320]	2007	$\mathcal{C},\mathcal{E},\mathcal{R}$	13	11		NN	5 arousal levels	31 / 62%
							5 valence levels	26 / 57%

source	year	signals	# subjects	features selection / reduction		classifiers	target	classification
[733]	2007	$\mathcal{C}$,$\mathcal{E}$,$\mathcal{R}$,$\mathcal{M}$	1	193	BPSO	k-NN	4 emotions	86%
[336]	2007	$\mathcal{C}$,$\mathcal{E}$,$\mathcal{R}$,$\mathcal{M}$,$\mathcal{T}$	3	77	SBS	k-NN,NN,LDA	4 emotions	51%–71%
[409]	2007	$\mathcal{C}$,$\mathcal{E}$,$\mathcal{M}$	3	54		SVM	3× 2 levels	85%/80%/84%
[697]	2007	$\mathcal{C}$,$\mathcal{E}$	40	28		reg. model	5 emotions	63%-64%
[542]	2007	$\mathcal{C}$,$\mathcal{E}$,$\mathcal{M}$,$\mathcal{T}$	5	18		FL,RT	anxiety scale?	57%-95%
[328]	2007	$\mathcal{E}$,other	24	14		k-NN,SVM,GP	2 frustrations	79.1%
[287]	2007	$\mathcal{C}$,$\mathcal{E}$,$\mathcal{R}$,$\mathcal{M}$,$\mathcal{T}$	24	4×50		LDA,GMM	2 levels of stress	94%-89%
[360]	2007	$\mathcal{C}$,$\mathcal{E}$,$\mathcal{R}$,$\mathcal{M}$,$\mathcal{O}$	34	23	ANOVA	PDA	3 emotions	69%-85%
[413]	2008	$\mathcal{C}$,$\mathcal{E}$,$\mathcal{M}$,$\mathcal{T}$	6	35		SVM	3 affect states	83%
[330]	2008	$\mathcal{C}$,$\mathcal{E}$,$\mathcal{R}$,$\mathcal{M}$	10	15		SVM,ANFIS	4 affect states	79%
[734]	2008	$\mathcal{C}$,$\mathcal{E}$	72	20	ANOVA	SVM,NN	2 fun levels	70%
[338]	2008	$\mathcal{C}$,$\mathcal{E}$,$\mathcal{R}$,$\mathcal{M}$	3	110	SBS	LDA, DC	4 emotions	70 / 95%
[102]	2008	$\mathcal{M}$	1	14	DWT	NN	4 emotions	75%
[397]	2008	$\mathcal{C}$,$\mathcal{E}$,$\mathcal{R}$,$\mathcal{M}$,$\mathcal{T}$	40	5		SVM	5 emotions	47%
							2 levels of arousal	82%
							2 levels of valence	72%
[410]	2008	$\mathcal{C}$,$\mathcal{E}$,$\mathcal{M}$	6	54		SVM	3×2 levels	81%
							3 levels	72%
[119]	2008	$\mathcal{C}$,$\mathcal{E}$,$\mathcal{M}$,$\mathcal{T}$	6	?		SVM,QV-learning	3× 2 levels	83%
			6	?		SVM	3 behaviors	81%
[102]	2008	$\mathcal{M}$	1	12	DWT	NN,TM	4 emotions	75%
[604]	2008	$\mathcal{C}$,$\mathcal{E}$	1	23	LDA	k-NN,SVM	4 emotions	83%-67%
		$\mathcal{C}$,$\mathcal{E}$,$\mathcal{B}$		53			4 emotions	87%
[41]	2008	$\mathcal{C}$,$\mathcal{E}$,$\mathcal{R}$,$\mathcal{T}$	1	225	SFS,LDA	LDA,k-NN,NN	4 emotions	90%
[95]	2009	$\mathcal{C}$,$\mathcal{E}$,$\mathcal{R}$	10	18		LDA,SVM	3 emotions	51%

source	year	signals	# subjects	features	selection / reduction	classifiers	target	classification
							2 emotions	66%
[97]	2009	$\mathcal{C}$,$\mathcal{E}$,$\mathcal{T}$	6	3		NN	4 emotions	88%
[289]	2010	$\mathcal{C}$,$\mathcal{E}$,$\mathcal{R}$,$\mathcal{B}$	15	38		SVM	2 arousal	77%
			15	21	t-test,LDA	LDA,SVM	2 valence	85%
[474]	2010	$\mathcal{C}$,$\mathcal{E}$,$\mathcal{T}$	34	12		k-NN,BP	3 emotions	65%–83%
[675]	2010	$\mathcal{E}$,$\mathcal{M}$	21	10	ANOVA,PCA	k-NN,SVM,NN	4 emotions	61%
[329]	2011	$\mathcal{C}$,$\mathcal{E}$,$\mathcal{M}$,$\mathcal{R}$	10	41		SVM,ANFIS,NB,	4 emotions	72%
						k-NN,TAN,RT	3 emotions	77%
							2 emotions	82%
[356]	2011	$\mathcal{C}$,$\mathcal{E}$,$\mathcal{R}$,$\mathcal{T}$, $\mathcal{M}$,other	34	14	SBS,SFS	LDA,QuadC,NN, k-NN	3 emotions	78/82%

Signals: $\mathcal{C}$: cardiovascular activity; $\mathcal{E}$: electrodermal activity; $\mathcal{R}$: respiration; $\mathcal{M}$: electromyogram; $\mathcal{B}$: electroencephalogram; $\mathcal{T}$: skin temperature; $\mathcal{P}$: pupil diameter; $\mathcal{O}$: Expiratory pCO_2

Classifiers: HMM: Hidden Markov Model; RT: Regression Tree; BN: Bayesian Network; NN: Artificial Neural Network; SVM: Support Vector Machine; LDA: Fisher's Linear Discriminant Analysis; k-NN: k-Nearest Neighbors; TAN: Tree Augmented Naïve Bayesian classifier; ANFIS: Adaptive Neuro-Fuzzy Inference System; and QuadC: Quadratic classifier; LinC; Linear classifier; FL: Fuzzy Logic System; TM: Template Matching classifier; GMM: Gaussian Mixture Model; BP: BackPropagation algorithm; and GP: Gaussian Process classification

Selection: DBI: Davies-Bouldin Index; PCA: Principal Component Analysis; reg.: regression; SFS: Sequential Forward Selection; SBS: Sequential Backward Selection; SFFS: Sequential Floating Forward Search; ANOVA: Analysis of Variance; and DC: Dichotomous Classification; BPSO: Binary Particle Swarm Optimization; B-W: Baum-Welch re-estimation; FP: Fisher Projection; and DWT: Discrete Wavelet Transform.

Note. * The authors used thresholds of 10% and 20%, which they denote as bandwidth. However, no definition was provided for this bandwidth. Hence, a comparison with this work is impossible.

2.4.2 Time for a change

Taken together, implicit messages of emotion are expressed through bodily (e.g., movements) and facial expressions [131, 192, 511, 652, 739] and by way of speech signal characteristics (e.g., intonation) [131, 182, 511, 590, 739]. In line with Picard [521, 524], I pose that this duo is not complete and physiological responses should be added to it to complete the pallet of affective signals. Although, such responses are hard to notice by humans, as is the case with various facial muscles [643]. In contrast, computing devices augmented by biosensors can record such signals, as has been shown in the last decade of research; see Table 2.4.

Biosignals have one significant advantage compared to visual, movement, and speech signals, they are free from social masking [643]. This is in sharp contrast to visual appearance and speech, which can all be (conveniently) manipulated to some extent [643], in particular by trained individuals such as actors. Moreover, an important advantage of biosignals over either speech or vision is that you get a continuous signal, as opposed to speech that is only of use when the person is speaking or facial expressions that tend to be sparse when people are doing, for example, computer work. So, biosignals enable communication, where traditional channels (i.e., vision and speech [148, 184]) are absent or fail (cf. [617]). So, par excellence, biosignals can augment HCI as well as human-human interaction [315].

To bring biosignals as affective signals from research to practice, however, significant improvements are needed. Although it is very possible that some closed loop applications function satisfactorily in practice, in general either the number of emotional states recognized is rather limited (often 2 to 4) or the ultimate classification accuracy is relatively low (often below 80%). So, there is significant room and need for improvement to obtain the high accuracy levels for the classification of multiple emotional states, which is necessary for the construction of smooth affective closed loops.

In the next three parts of this monograph, Parts II, III, and IV, I set out a series of studies to systematically review the options for improvement that are still open for *ASP*. These studies all address issues that are crucial for the development of closed loop *ASP*, as presented in Section 1.5. In particular, they are of importance for its signal processing + pattern recognition pipeline. These three parts will be succeeded by an epilogue in which the first chapter presents guidelines for each of these two steps in the processing pipeline. This monograph will now first continue with two chapters that employ four biosignals (i.e., $3\times$ EMG and EDA), uses dynamic stimuli (i.e., movie fragments) to induce emotions, and explores the importance of the length of time windows for *ASP*.

II. Baseline-free ASP

3

Statistical moments as signal features

Abstract

To improve Human-Computer Interaction (HCI), computers need to be able to recognize and respond properly to their user's emotional state. This is a fundamental application of *affective computing*, which relates to, arises from, or deliberately influences emotion. As a first step to a system that recognize emotions of individual users, this research focused on how emotional experiences are expressed in six parameters (i.e., mean, absolute deviation, standard deviation, variance, skewness, and kurtosis) of not baseline-corrected physiological measurements of the ElectroDermal Activity (EDA) and of three ElectroMyoGraphy (EMG) signals: frontalis (EMG1), corrugator supercilii (EMG2), and zygomaticus major (EMG3). Twenty-four participants were asked to watch film scenes of 120 seconds, which they then rated. These ratings enabled us to distinguish four classes of emotions: negative, positive, mixed, and neutral. The skewness and kurtosis of the EDA, the skewness of the EMG2, and four parameters of EMG3, discriminate between the four emotion classes and explained $36.8\% - 61.8\%$ of the variance between the emotion four classes. This, despite the coarse time windows that were used. Moreover, rapid processing of the signals proved to be possible. This enables tailored HCI facilitated by an emotional awareness of systems.

This chapter is an adapted and extended version of:

Broek, E.L. van den, Schut, M.H., Westerink, J.H.D.M., Herk, J. van, and Tuinenbreijer, K. (2006). Computing emotion awareness through facial electromyography. *Lecture Notes in Computer Science (Human-Computer Interaction), 3979*, 51–62.

which is also published as:

Westerink, J.H.D.M., Broek, E.L. van den, Schut, M.H., Herk, J. van, and Tuinenbreijer, K. (2008). Computing emotion awareness through galvanic skin response and facial electromyography. In J.H.D.M. Westerink, M. Ouwerkerk, T. Overbeek, W.F. Pasveer, and B. de Ruyter (Eds.), *Probing Experience: From Academic Research to Commercial Propositions (Part II: Probing in order to feed back), Chapter 14*, p. 137–150. Series: Philips Research Book Series , Vol. 8. Dordrecht, The Netherlands: Springer Science + Business Media B.V.

and is filed as:

Westerink, J.H.D.M., Broek, E.L. van den, Schut, M.H., Tuinenbreijer, K. (2007). Higher order GSR-measurement interpretation indicating emotions. International Patent Application No. PCT/IB2008/050477 (PH007322), filed on February 11.

3.1 Introduction

Computers are experienced by their users as cold hearted (i.e., "*marked by lack of sympathy, interest, or sensitivity*" [448]). However, 'during the past decade rapid advances in spoken language technology, natural language processing, dialog modeling, multi-modal interfaces, animated character design, and mobile applications all have stimulated interest in a new class of conversational interfaces' [504]. The progress made in this broad range of research and technology enables the rapid computation and modeling of empathy for human-computer interaction (HCI) purposes. The latter is of importance since conversation is, apart from being an information exchange, a social activity, which is inherently enforcing [504]. Futurists envision embodied, social artificial systems that interact in a natural manner with us. Such systems need to sense its user's emotional state.

Empathic artificial systems can, for example, prevent user frustration in HCI. Users frequently feel frustrated by various causes; for example, error messages, timed out/dropped/ refused connections, freezes, long download time, and missing/ hard-to-find features [94]. Picard [518] posed the prevention of user frustration as one of the main goals in HCI. When prevention is not sufficient, online detection and reduction of frustration is needed. Biosignals are useful in detecting frustration [521]. According to Hone [286], an (embodied) affective agent, using techniques of active listening and emotion-awareness could reduce user frustration.

The current chapter discusses the emotions people can experience and their expression in and detection through *ASP*, in Section 3.2 and Section 3.3. Next, in Section 3.4, affective wearables are introduced in which the proposed apparatus for the measurement of the biosignals can be embedded. In Section 3.5, we present an experiment into the appropriateness of various statistical measures derived from biosignals, followed by a reduction of the data in Section 3.6. The experimental results are described in Section 3.7. The chapter ends with Section 3.8 in which the results are discussed, limitations are denoted, and future research is described.

3.2 Emotion

Despite the complexity of the concept of emotion, most researchers agree that emotions are acute affective states that exist for a relatively short period of time and are related to a particular event, object, or action [502, 521]. In relation with physiology, emotions are predominantly described as points in a two-dimensional space of affective valence and arousal, in which valence represents overall pleasantness of emotional experiences ranging from negative to positive, while arousal represents the intensity level of emotion, ranging from calm to excited [372, 647]. This allows us to tell the difference between 4 rough classes of emotions, when differentiated between both high and low valence and high and low arousal. Some researchers even differentiate between nine classes by including a neutral section on both the valence and arousal axes. However, in principle, any number of classes can be defined, where the valence and arousal axes are not necessarily divided with the same precision [61].

The valence-arousal model, however, does not account for mixed emotions: positive and negative at the same moment. In order to be able to cope with mixed emotions, Larsen et al. [380] and Konijn and Hoorn [357] suggest that valence should be unipolar instead of bipolar. When valence is rated on two scales, one for the intensity of positive affect and one for the intensity of negative affect, mixed emotions, in the sense of both positive and negative emotions, will show. As an extension to the valence-arousal model, a unipolar valence axis, with separated positive and negative axes, might allow for a better discrimination between different emotions.

In the current research, we only explored the valence axis. The reason is that the simplest differentiation of emotions is a differentiation between positive and negative emotions. In most cases of HCI, this is sufficient to improve the dialog between user and computer; for example, when a user has a negative emotion, the computer can adapt its dialog to that, depending on the context.

3.3 Measures of affect

The roots in research toward psychophysiological aspects of emotions lay in Darwin's book '*The expression of emotions in man and animals*', which he wrote in 1872. The overall assumption is that emotion arouses the autonomic nervous system (ANS), which alters the physiological state. This is expressed in various physiological measures, often stimulated through the ANS; for example, heart rate, blood pressure, respiration rate, ElectroDermal Activity (EDA), and muscle activity (see Table 1.1). The main advantage of using autonomic physiological measures is that autonomic variables are regulated by the ANS, which controls functions outside the individual's conscious control [85]. In this research, we focused on how emotional experiences, rated to their positive and negative affect, are expressed in four biosignals:

- EDA (also termed GSR) [62], which is a measure of the conductivity of the skin: arousal of the ANS influences sweat glands to produce more sweat; consequently, skin conductivity increases. EDA was chosen because it is an autonomic variable; hence, it cannot be controlled by the user [136].
- Three EMG signals: frontalis, corrugator supercilii, and zygomaticus major [664]. EMG measures muscle activity of a certain muscle. These measures were chosen because a great deal of emotional expression is located in the face [380, 592, 664]. Facial EMG

is related to affective valence; however, the type of relation depends strongly on the muscle that is measured [133, Chapter 10], [85, Chapter 12], [396, Chapter 9, 1st ed.; Chapter 11]. The corrugator supercilii, which causes a frown when activated, increases linearly with a decrease in valence, while the zygomaticus major, which is responsible for smiling when activated, increases with an increase in valence [373, 592, 664]. The EMGs of these two muscles are known to discriminate best between emotions; see [396, Chapter 9, 1st ed.] for a concise review. The frontalis is a measure for attention (and fatigue) and, hence, is not expected to discriminate between emotions [133, Chapter 10].

These measures have extensively proven their use to detect emotional experiences in laboratory settings, mostly in group-averaged, baseline-corrected paradigms. In order to make them useful for emotion-aware systems, three aspects will have to change:

1. the measurements will have to be done in a less obtrusive manner;
2. the interpretation of the signals will have to be meaningful on an individual (not a group-averaged) level; and
3. robust signal interpretation algorithms will have to be developed that are baseline-free or incorporate automatic (non-manual) baseline correction.

The first issue will be dealt with in the next paragraph, where we will discuss the advent of unobtrusive affective wearables. Our focus for the remainder of the chapter is on the search for robust signal interpretation algorithms that do not need a manual baseline correction.

3.4 Affective wearables

Using the EDA and EMG signals, a system will be able to determine the emotional state of its user, certainly if that system also possesses a user-profile. Affective wearables will facilitate such a system in monitoring the user in an unobtrusive manner. Direct physiological measures are often considered to be obtrusive to the user, but this is not necessarily true. In the field of *affective computing*, some efforts have been made to design unobtrusive measurement technology: affective wearables. Picard [521] defines an affective wearable as "*a wearable system equipped with sensors and tools which enables recognition of its wearer's affective patterns*". Affective wearables will become smaller in time, due to improved design and smaller technology components. Affective wearables could make a huge difference in user acceptance of direct physiological measures, especially when hidden in daily used tools and objects.

The acceptance of direct physiological measurements is of great importance since indirect physiological measurements are much more subject to noise. Indirect physiological measurements (e.g., through speech analysis [677]) have been applied in controlled settings such as telepsychiatry [279] and evaluation of therapy effectiveness [677]. However, outside such controlled conditions these measures have not proven to be reliable.

Measurement of biosignals have already been embedded into wearable tools; for example, Picard and Scheirer [523] designed the 'Galvactivator', a glove that detects the skin conductivity and maps its values into a LED display. In an overview of previous work of the *Affective computing* Research Group at MIT, Picard [522] describes several affective wearables. One affective wearable that is of interest in this research is a pair of expression glasses. The pair of expression glasses sense facial movements, which are recognized as affective patterns.

3.5 Experiment

The goal of this experiment was to enable a search for robust (e.g., baseline-free) algorithms for use in future emotional awareness systems. These should interpret positive or negative emotions from biosignals.

3.5.1 Participants

In the experiment, 24 Dutch subjects participated (average age: 43 years). Twenty of the participants were females, since we expected clearer facial emotion expressions from them [66, 361]. As we could not find 4 more females, we replaced them by men in order to be able to maintain the counterbalancing in the experiment design, as will be depicted in Section 3.5.3. All subjects had been invited from a volunteer subjects database, and were rewarded with a small gift for their participation. All subjects signed an informed consent form.

3.5.2 Equipment and materials

We selected 16 film fragments for their emotional content. Most were adopted from the set of Gross and Levenson [235, 237] and are known to elicit one unique emotion from various viewers: Silence of the Lambs (198 seconds), When Harry met Sally (149 seconds), The Champ (153 seconds), Sea of Love (9 seconds), Cry Freedom (142 seconds), The Shining (80 seconds), Pink Flamingoes (30 seconds). We used these fragments in English-spoken versions with Dutch subtitles, as is usual on Dutch TV and in Dutch cinemas. Since we were not able to find enough material of Gross and Levenson [235, 237] with Dutch subtitles of acceptable quality, we added a number of similar fragments to the set: Jackass the Movie -

paper-cut scene (51 seconds), Static TV color bars (120 seconds), The Bear - intro (120 seconds), Sweet Home Alabama - wedding scene (121 seconds), Tarzan - orchestra scene (133 seconds), Abstract Shapes - screen saver (120 seconds), Lion King - dad's dead (117 seconds); Nature documentary (120 seconds), Final Destination - side-walk café scene (52 seconds). An overview of all film fragments used in this study is provided in Table 3.1. The duration of the 16 film fragments ranged from 9 seconds to 4 minutes. For the fragments with durations shorter than 120 seconds, a plain blue screen was added to make a total of 120 seconds, a minimum duration needed for assessing both the low and high frequency HRV components [44]. We displayed the film fragments on a large $42''$ $16:9$ flat panel screen attached to the wall. The subjects viewed the fragments from a comfortable chair at a distance of about 2 meters.

We used a TMS International Porti$5-16$/ASD system for the psychophysiological measurements. The system was connected to a computer with TMS Portilab software*. Its ground electrode was attached to the subject's right-hand side lower chest area. We performed 3 EMG measurements: at the right-hand corrugator supercilii muscle, the left-hand zygomaticus major muscle, and the frontalis muscle above the left eye (see Figure 3.1). The detail muscle positions were found by touching the contracting muscles, in line with

*URL TMS Portilab software: http://www.tmsi.com/

Table 3.1: The eight film scenes with the average ratings with the accompanying standard deviations (between brackets) given by subjects ($n = 24$) on both experienced negative and positive feelings. Four emotion classes are founded: neutral, mixed, positive, and negative, based on the latter two dimensions. The top eight film scenes were selected for further analysis.

Film scene	Positive	Negative	Emotion category
Color bars	1.60 (1.43)	2.20 (2.04)	neutral
Abstract figures	1.20 (0.70)	2.10 (1.94)	neutral
The bear	5.15 (1.50)	1.65 (0.88)	positive
Tarzan	5.10 (1.17)	1.50 (0.95)	positive
Final destination	3.11 (1.70)	4.32 (1.63)	mixed
Lion King	3.85 (2.21)	3.65 (1.93)	mixed
Cry freedom	1.95 (1.54)	6.25 (1.07)	negative
Pink flamingos	1.75 (1.20)	5.60 (1.54)	negative
Silence of the lambs	2.30 (1.38)	3.85 (1.73)	neutral
When Harry met Sally	4.60 (1.47)	1.80 (1.15)	positive
The champ	2.65 (1.46)	4.35 (1.05)	mixed
Jackass the movie	1.85 (1.57)	5.95 (1.47)	negative
Sea of love	2.15 (1.31)	3.90 (1.74)	neutral
Sweet home Alabama	4.35 (1.66)	1.70 (1.26)	positive
The shining	2.65 (1.39)	3.55 (1.47)	neutral
Nature documentary	4.50 (2.04)	1.45 (1.28)	positive

Lapatki, Stegeman, and Jonas' [377] recommendations. Subsequently, for each measurement we placed 2 electrodes along the muscle (see Figure 3.1), respecting Fridlund and Cacioppo's [206] *"Guidelines for human electromyographic research"*. The EMG signals were first high-pass filtered at 20 Hz; then, the signal was rectified by taking the absolute difference of the two electrodes and finally a central moving average filter was applied with a time constant of 0.2 seconds.

Two active skin conductivity electrodes were attached to the subject's right hand: on the inside distal phalanges of the index and ring fingers (see Figure 3.1). We calculated skin conductivity from the measured signal by central moving average filtering with a time constant of about 2 seconds; thus, capturing EDA signal variations reliably in first order [62]; see also Table 1.1 in Chapter 1. ECG was also measured with the intention of investigating heart rate variability measures, but since the TMS program failed to actually record the data for many participants, these data were not analyzed.

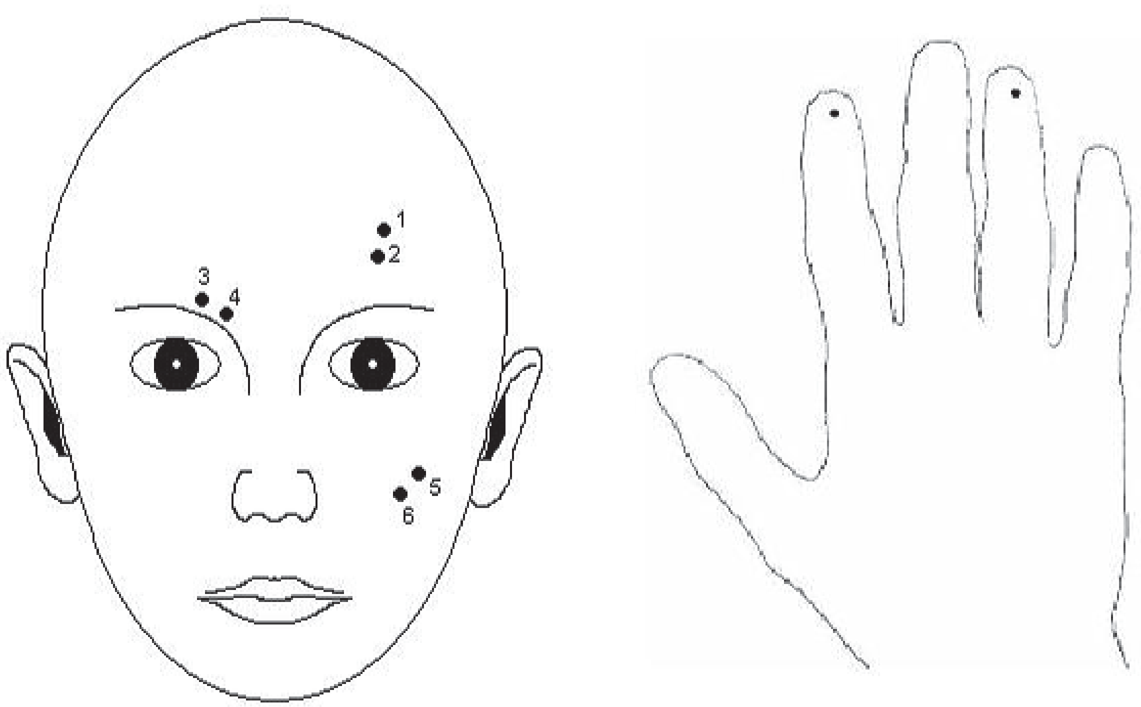


Figure 3.1: Left: The points indicate the electrodes that were placed on the face of the participants to determine the EMG signals. The EMG signals of the frontalis, corrugator supercilii, and zygomaticus major were respectively measured through electrodes $1-2$, $3-4$, and $5-6$. Right: The points at which the electrodes were placed on the hands of the participants to determine the EDA signal.

3.5.3 Procedure

After the subject was seated, the electrodes were attached to their chest, their fingers, and their face. Then, we checked the recording equipment and adjusted it if necessary. After a 5-minute rest period, the 16 video fragments were presented to the subject in pseudo-random order, so that positive and negative scenes were spread evenly over the session. Twelve subjects received that same pseudo-random order, though each started with a different scene in the list. The remaining 12 subjects were given the reverse pseudo-random order, again each starting with a different scene. We presented a plain blue screen for 120 seconds between two fragments, to allow the effects of the previous film fragment to fade out.

The entire viewing session lasted slightly over one hour, after which we removed the electrodes. Next, the subjects were asked to answer a few questions regarding each of the film fragments viewed. We deliberately did not ask these questions directly after each individual film fragment, since this would direct the participants' attention to the questioned items in all subsequent viewing, which would have given the rest of the viewing session an unnatural character. In order to help them recall their feelings during the presentation of the film fragments, the participants were sequentially provided with representative print-outs of each fragment. For each film fragment, they were asked to rate, on a 7-point Likert scale, the intensity of positive feelings they had had while watching it, as well as the intensity of negative feelings, and the amount of arousal. With these three axes we expected to include the both axes of Russel's valence-arousal model [372, 566, 647], as well as the possibility of mixed emotions [79, 92, 357, 379]. As we needed to present separate scales for positive and negative feelings in order to capture possible mixed emotions, we could not deploy the Self Assessment Mannequin (SAM) [372].

3.6 Data reduction

For each video fragment, we calculated the average positive rating as well as the average negative rating. Based on these averages, we could classify the fragments into 4 emotion classes: neutral, mixed, positive, and negative. In order to obtain an even distribution over emotion classes, we selected two fragments in each emotion category for further analysis. In each category, we chose the fragments with a duration closest to 120 seconds, so that time effects could more easily be compared (see Table 3.1). This resulted in the following set for further analysis: Color Bars and Abstract Figures (both 'neutral', with both ratings below 2.5), The Bear and Tarzan (both 'positive', with positive ratings above 5.0 and negative ratings below 2.0), Final Destination and Lion King (both 'mixed', with both positive and negative ratings above 3.0), and Cry Freedom and Pink Flamingoes (both 'negative', with negative ratings above 5.0 and positive ratings below 2.0).

Not all biosignals data were fit for analysis: the EMG signals of 2 subjects were corrupted, probably due to loose contacts, and we decided not to include these data sets in further analyses. Moreover, for the same reason, the recordings of one subject, during the film scene of the "Pink flamingos", were skipped. For the remaining 22 subjects, we processed the 4 biosignals to obtain the following measures: mean, absolute deviation, standard deviation, variance, skewness, and kurtosis.

Mean, absolute deviation and standard deviation are well-known dimensional quantities with the same units as the measured signal. Variance is also a frequently used parameter. The skewness and kurtosis, however, are expressed as non-dimensional quantities; see [197] for their introduction. [318] provide a comprehensive overview and a comparison of the skewness and kurtosis measures for both normal and non-normal distributed samples. In this overview, they state that it is suggested that "*skewness and kurtosis should be viewed as 'vague concepts', which can be formalized in many ways. Accordingly, many different definitions have been proposed.*" For this research, we adopted the following descriptions: Skewness characterizes the degree of asymmetry of a distribution around its mean and kurtosis characterizes the relative peakedness and tail weight of a distribution.

Following the literature [81, 318, 534, 710], we define skewness and kurtosis for samples $\{x_1, x_2, \ldots, x_N\}$ as:

$$\mathrm{Skewness}(x_1, x_2, \ldots, x_N) = \frac{1}{N} \sum_{j=1}^{N} \left[\frac{x_j - \overline{x}}{\sigma} \right]^3 \tag{3.1}$$

and

$$\mathrm{Kurtosis}(x_1, x_2, \ldots, x_N) = \frac{1}{N} \sum_{j=1}^{N} \left[\frac{x_j - \overline{x}}{\sigma} \right]^4 - 3 \tag{3.2}$$

with σ being the standard deviation and $\overline{x}$ being the mean of the data set. For a normal distribution, the third and fourth central moments are respectively 0 and 3 [197, 318]. Since our objective was to describe both skewness and kurtosis relative to that of a normal distribution, a correction of -3 was applied for kurtosis, as is often done.

3.7 Results

For each of the six statistical parameters, for each film fragment, and for each subject, the complete EDA and EMG signals were processed over the last 120 seconds of the film fragment. The duration of 120 seconds was chosen because it was available for the majority of the scenes. Two film fragments were shorter than that, and for them we included measurements taken during the blue screen following it in order to add up to a section of 120 seconds as well (see also Section 3.5.2). Note that we deliberately did not correct these values for their baseline, because - although useful in academic research - the baseline-correction procedure is not easily applicable in future emotionally aware systems.

For each parameter of each physiological measure, a repeated measures ANOVA was conducted, with the four emotions, each measured with two film scenes, as within-subject factors. So, a total of 24 (i.e., 4×6) repeated measures ANOVAs were conducted. As measure of effect size partial eta squared (η^2) is reported, which indicates the proportion of variance accounted for (i.e., a generalization of r/r^2 and R/R^2 in correlation/regression analysis) [211, 737]. The classification of the film scenes into the four emotion classes was based on the participants' ratings as provided in Table 3.1, which were perfectly in line with the findings reported by Gross and Levenson [235, 237].

The EMG of the frontalis did not provide a significant discrimination between the 4 emotion classes on any of the statistical parameters. Of all physiological measures, the zygomaticus major signal is the most discriminative biosignal (see Table 3.2). The mean, absolute deviation, standard deviation and variance calculated over the zygomaticus major EMG signal showed strong significant effects of emotions. Significant effects did also show in the skewness and kurtosis of the EDA signal and the skewness of the corrugator supercilii EMG signal (Table 3.2). For the skewness of the EMG zygomaticus signal a trend was present over the four emotions ($F(3, 18) = 3.013, p = .057$), which explained rather a lot of the variance present between the 4 emotion classes ($\eta^2 = .334$).

3.8 Discussion

3.8.1 Comparison with the literature

Most 120 seconds averaged values of the biosignals yielded no significant effects of emotion class, in contrast to what is generally reported in the literature. One of the reasons might be that we chose not to correct our data for baseline values, as is common in the psychophysiological literature. Another factor is that the present analysis was chosen to extend over a relatively long period of time including the beginning of the video fragment in which the targeted emotions were still in the process of being elicited, which might have diminished the differences between categories of emotions.

For the zygomaticus major, we did find an effect for the average value, even when not corrected for baseline and averaged over 120 seconds. This is in line with results of previous research by Larsen, Norris, and Cacioppo [380], who concluded that valence influences both the corrugator supercilii and the zygomaticus major. They found that valence had a stronger effect on the corrugator supercilii than on the zygomaticus major in experiencing

Table 3.2: The discriminating statistical parameters for the EDA, EMG corrugator supercilii, and EMG zygomaticus signals. For each parameter, the average value for all four emotion classes (i.e., neutral: 0; positive: +; mixed: +/-; negative: -.) is provided as well as the strength and significance of its discriminating ability. Additionally, as measure of effect size partial eta squared (η^2) is reported, which indicates the proportion of variance accounted for [211, 737].

Physiological measure	Statistic parameter	average value on				effect		
		-	+	+/-	-	$F(3,18)$	p	η^2
EDA	skewness	0.46	0.01	-0.15	0.39	7.289	$= .002$	.549
	kurtosis	-0.66	-0.78	0.55	-0.19	3.812	$= .028$	.388
EMG frontalis	–							
EMG corrugator supercilii	skewness	1.99	2.84	3.49	3.29	3.500	$= .037$	.368
EMG zygomaticus	mean	2.74	5.21	3.15	3.53	9.711	$< .001$	.618
	abs. dev.	1.64	3.77	2.10	2.42	8.369	$< .001$	.583
	SD	2.46	6.01	3.68	3.96	5.837	$= .006$	.493
	variance	7.23	63.82	18.69	23.21	4.064	$= .023$	.404

standardized affective pictures, sounds, and words, while our research shows a stronger effect of the four emotion classes on the mean zygomaticus major signal, than on the corrugator supercilii. In addition, the effect is present with four statistical parameters of the zygomaticus major, where it is only present in one statistical parameter (skewness) of the corrugator supercilii.

The difference in strength of the effects found between the current research and that of Larsen, Norris, and Cacioppo [380] can possibly be explained by the absence of a baseline correction in our procedure. Another difference between the two researches is the type of stimuli (cf. [8]). Film scenes are dynamic and multi-modal, they induce emotions by both auditory, and dynamic visual stimuli, as well as affective words, in some fragments. The dynamic and multi-modal characteristics of the film scenes also provide good means to build up emotions, or to create a shock effect [570, 700, 701]. This is almost not possible with affective words, sounds or pictures of a static character, as their use lacks the opportunity to built up emotions. On the one hand, all these factors give film scenes a relatively high degree of ecological validity [235, 237, 700, 701]. On the other hand, it is not possible to determine which modality influences the emotional state of the subjects to the highest extent.

For three of the 4 biosignals the parameter skewness turned out to be important as a significant effect or as a trend. To the authors best knowledge, the skewness (and kurtosis) of EMG signals as discriminating descriptor have been discusses in only three studies. In 1983, Cacioppo, Marshall-Goodell and Dorfman [82] analyzed among a number of parameters, the skewness and kurtosis of skeletal muscle patterns, recorded through EMGs. Four years later, an article by Cacioppo and Dorfman [81] that discussed "*waveform moment* *analysis in psychophysiological research*" in general. In 1989, Hess et al. [278] conducted research toward experiencing and showing happy feelings, also using video segments. Hess et al. [278] recorded four facial EMG signals and extracted the mean, variance, skewness and kurtosis of these signals. The current research is distinct from that of Hess et al. [278] since it distinguishes four emotion classes instead of the presence or absence of only one. Each of these three studies identified skewness and kurtosis of EMG signals as potentially interesting for the discrimination between emotions. However, surprisingly little attention has been given to moments of order 3 and higher in *ASP*.

3.8.2 Use in products

Not all investigated parameters of all measures proved to be equally suited for sensing human's emotional states. This is no doubt due to the demanding analysis conditions we imposed: no baseline correction and averages over relatively long time intervals. Nevertheless, even under these demanding analysis conditions, some of the measures still succeeded in distinguishing between the respective emotion classes.

For three of the four biosignals used, the parameter skewness proved to be an interesting source of information. The skewness of the distributions of the data of two of the biosignals differs significantly over the four emotions, where a trend is present for a third signal. To inspect more distribution details of the signals, additional analyses could be conducted. Measures such as the slope of the signal and the peak density could be taken into account for further analysis. Such analysis could help understanding to what extent emotions were indeed built up during the movie scenes.

In addition to adding more descriptors of the biosignals, the time windows of measurement can be changed. In the current setup, the time window enclosed the complete length of the film scene. However, smaller time windows (e.g., 10 or 30 seconds) can be applied to conduct more detailed analysis of biosignals' behavior in relation to the movie content. Moreover, dynamic time windows can be applied that enclose the time directly after a critical event (if any) appeared in the film scene. The drawback of the latter approach is that it cannot be applied in practice, while it may be expected to provide good results for data gathered through experimentation, as in the current research.

A more general notion that can have a significant impact on measurement of emotions is that the emotional state of people changes over time, due to various circumstances. Moreover, different persons have different emotional experiences over the same events, objects, or actions. This variance in experienced emotions is determined by a person's personality. Personality traits correlate with affective states, especially with the personality traits extraversion and neuroticism, which have been linked both theoretically and empirically to the fundamental affective states of positive and negative affect, respectively [442]. Hence, to enable tailored communication strategies in HCI, not only the emotional state of a person should be determined but also his personality. When the system possesses a personality profile of its user, it will be able to react appropriately to its user's emotions by selecting a suitable communication strategy. We will explore this issue in Chapters 5 and 6.

The next chapter will continue the analyses presented in this chapter. Analysis will be conducted on the same data set using other time windows. Events in the movie fragments will be traced and their effects on the EMG and EDA signals will be unveiled. Moreover, the possible influence of scene changes will be addressed.

4

Time windows and event-related responses

Abstract

Emotion aware consumer products require reliable (i.e., unobtrusive, robust, and lacking calibration) short-term emotion assessment. To explore the feasibility of this, the data presented in the previous chapter was analyzed again. The unfiltered biosignals were processed and six statistical parameters (i.e., mean, absolute deviation, standard deviation, variance, skewness, and kurtosis) were derived for each 10-sec interval of the film fragment. For each biosignal, skewness and kurtosis discriminated between affective states, accompanied by other parameters, depending on the signal. The skewness parameter was also shown to indicate mixed emotions. Moreover, a mapping of events in the fragments on the signals showed the importance of short-term emotion assessment. Hence, this research identified generic features, denoted important considerations, and illustrated the feasibility of emotion-aware consumer products.

This chapter is a compressed version of:

Broek, E.L. van den & Westerink, J.H.D.M. (2009). Considerations for emotion-aware consumer products. *Applied Ergonomics, 40(6)*, 1055–1064. *[Special issue: Psychophysiology in Ergonomics]*

4.1 Introduction

There is a growing interest in systems that are aware of user's emotions. Such systems find their domain of application in professional or specialized applications, such as emotional support for people with autism [646], stress in ambulance dispatchers [467], irritation detection to support call center employees [156], as therapy progress indicator for psychologists [677], or with pilots and airline crews to determine their arousal [657]. In a typical consumer context, however, there is no explicit task at hand and the main intention is to support a pleasant every-day life. In such a context, emotion-aware systems can adapt the conversational dialogue in order to optimize HCI, can characterize someone's emotional state for them for increased self-awareness or for others for enhanced communication, or they can adapt the user's environment to the present mood.

The consumer context poses a number of boundary conditions that might be different with respect to those of the professional context [63, 99, 722]. A first distinction is in the accuracy required for emotion detection. Though any consumer or professional application would preferably comprise a flawless emotion awareness system, it is likely that every now and then the emotion detection will be incorrect. It is to be expected that such errors are more detrimental in a professional application than in a consumer application, since they interfere with the professional task. In many consumer applications, such a task is often less prominent or even absent, and the system's reactions are not rigidly classified as correct or wrong, but rather as more or less preferred. Thus, a consumer application is somewhat more resilient with respect to emotion misclassifications, and most probably a higher percentage of misclassifications will be acceptable.

A second point of difference pertains to the unobtrusiveness of the application. If one wears a system either for professional use or to compensate for a certain handicap, one will more easily accept that the actual use of the system is not 100% comfortable [382, 693, 722]. For a consumer system, however, the emotion awareness system should preferably be unnoticeable to the user, the ultimate perceived ease of use. For instance, it could work from a distance, such as in speech or video processing. There, the detection of emotional features in the speech spectrum or in the facial expression can be done even without the awareness of the user; however, the physical range in which these systems work is limited. For a discussion on this issue, we refer to Chapter 2. To overcome this range of problems, the system could be worn. Then, it is important that it is not constantly noticeable to the user. Another form of obtrusiveness is when the system needs constant (re-)calibration. Where a professional application can require regular calibrations in order to improve the accuracy of the awareness classifications, this is not the case for typical consumer use. There, regular baseline-measurement periods or other explicit calibration actions interfere with the wish to live everyday-life without hassle. Thus, the algorithms employed for emotion classification should preferably be self-calibrating, especially in consumer-style applications.

A third issue in emotion aware systems, both for professionals and consumers, is time [382, 722]. Some applications work best when they can identify emotions over a relatively long period of time. For instance, when moods are being measured they are generally expected to last for hours, and change only gradually [567]. In contrast, in other applications emotions vary rapidly and a quick response of the system to changes in emotion is required. This is especially the case for applications in the realm of communication. There, emotions might come up quickly and it is important that the measured emotions are timely and adequately accommodated to facilitate a natural continuation of the conversation. Regardless of whether you want to convey your emotions privately to your partner or publicly through expressive clothes [37]; a broadly time-averaged signal will hide the intensities and will introduce interpretation delays. Also, if your home-computer or television set is to detect your frustrations and emotions, and react to it in a soothing way, this should be done immediately.

Thus, it appears that designing emotion-aware consumer applications is in fact much like striking a balance: for professional systems, some obtrusiveness can be accepted, provided that emotion classification is accurate and timely. For consumer-style systems, the balance is different: here unobtrusiveness and timely reactions are prime. But neither the drive to make such a consumer system unobtrusive, excluding calibrations, nor that to design the system to react quickly to emotion changes in small time intervals, will add to the accuracy of emotion detection. Luckily, there is some leeway: in consumer style systems, emotion classification errors are probably less detrimental. The hope is that the decline in accuracy is within acceptable limits. The research to be presented in the remainder of this chapter derives from this observation. We want to investigate whether it is possible to develop a method that captures quick changes with reasonable accuracy, but does not need a baseline correction or a personality profile and is noise-resistant.

The current chapter adopts the data presented in the previous chapter. The aim of the experiment with which this data was gathered was to check whether or not it is possible to capture quick emotional reactions with self-calibrating algorithms. As a consequence of these time and self-calibration requirements, we opted in the previous chapter for an experiment in which the subjects' emotions were elicited, using film fragments that are known to be powerful in eliciting emotions in laboratory settings [235, 237, 700, 701]. We chose biosignals that are commonly known to reflect emotions in the traditional (though not necessarily unobtrusive) way of baseline-corrected and broadly time-averaged signal processing to ensure that at least some emotion information was captured. The data that will be analyzed in this chapter has been adopted from an experiment that was already described in the previous chapter. Therefore, I will refrain from repeating the complete experimental setup and refer to Chapter 3 for specifications on this.

4.2 Data reduction

The classification of emotions is pictured integrated in various consumer applications. In a quest to find algorithms suitable for consumer contexts, a set of requirements for processing biosignals of affect can be specified, namely:

1. Short-term assessment; therefore, 10 second time windows are chosen.
2. Real-time processing; hence, as for the analyses in Chapter 3, baseline corrections are omitted from the processing scheme.
3. Robustness against small-scale measurement errors that last only a relatively short time interval; hence, distorted signals were not removed from the data set.
4. Good performance without personal profiles. At home and with some ubiquitous applications a personal profile can be easily included and will probably boost the performance of the emotion classification. However, for various consumer applications the use of such a profile cannot be realized; for example, in detecting customers' emotions. Hence, as for the analyses in Chapter 3, no personality characteristics were taken into account in processing and analyzing the signals.

Despite the experimental character of this research, the requirements mentioned above should be fully met. Thus, all 6 statistical measures have been calculated for all 4 biosignals for each 10-second interval of each of the 8 selected video fragments, which are the same as those selected for the analyses presented in Chapter 3. They were *neither* baseline-corrected *nor* cleaned up with respect to small-scale distortions, as these are generally manual, time-consuming operations, which are not plausible in a consumer context.

4.3 Results

With the first global analyses, as described in the previous chapter, we found that the newly introduced statistical parameters skewness and kurtosis had good discriminating abilities. In contrast, most averaged values of the biosignals did not yield significant effects. This is hardly surprising considering the coarse method of averaging over an interval of 120 seconds, ignoring typical events and changes in scenes. Therefore, a new series of analyses focussed on short-term emotion assessment, in twelve subsequent 10-sec time windows. The analysis of the data comprised three phases:

1. Determination of the possible influence of scene changes within the video fragments;
2. Analysis of the individual film fragments; and
3. Mapping of the events that occur in the fragments and the behavior of the biosignals at that moment.

4.3.1 The influence of scene changes

Each video fragment consists of a concatenated series of shots, generating abrupt changes at their transitions. The density of such scene changes might have a non-emotional impact on the viewer. Therefore, this section will describe the nonparametric correlations (Spearman's Rho, two-tailed) between the density of scene changes and the features of the biosignals. To this end, both the density of scene changes and the biosignals were determined for each time window of 10 seconds of each video fragment. The fragments color bars and abstract shapes were omitted from the analysis since they did not contain scene changes.

The Bear: The mean EDA ($r_s = .621, p = .031$), the absolute deviation ($r_s = .719, p = .008$) and variance ($r_s = .636, p = .026$) of the EMG frontalis, and the mean ($r_s = .716, p = .009$), absolute deviation ($r_s = .654, p = .021$), SD ($r_s = .737, p = .006$), variance ($r_s = .581, p = .047$), and kurtosis ($r_s = .610, p = .035$) of the EMG corrugator supercilii all correlated significantly.

Tarzan: The mean ($r_s = .642, p = .024$) and SD ($r_s = .528, p = .078$) of the EMG corrugator supercilii both correlated significantly.

Final Destination: The mean EDA ($r_s = .790, p = .002$), the skewness of the EMG frontalis ($r_s = .619, p = .032$), and the mean ($r_s = .638, p = .026$) and variance ($r_s = .871, p < .001$) of the EMG corrugator supercilii all correlated significantly.

Lion King: The kurtosis of the EMG frontalis ($r_s = .580, p = .048$) correlated.

Cry Freedom: The mean EDA ($r_s = .672, p = .017$), the skewness of the EMG frontalis ($r_s = .643, p = .024$), and the mean ($r_s = .665, p = .018$), absolute deviation ($r_s = .657, p = .020$), SD ($r_s = .643, p = .024$), and variance ($r_s = .621, p = .031$) of the EMG corrugator supercilii all correlated significantly.

Pink Flamingos: The mean EDA ($r_s = .776, p = .003$), the absolute deviation ($r_s = .726, p = .008$) and the SD ($r_s = .713, p = .009$) of the EMG frontalis, and the mean ($r_s = .651, p = .022$), variance ($r_s = .813, p = .001$), and skewness ($r_s = .713, p = .009$) of the EMG corrugator supercilii, and the absolute deviation ($r_s = .813, p = .001$), SD ($r_s = .813, p = .001$), and variance ($r_s = .776, p = .003$) all correlated significantly.

In the analysis described in the following three subsections, the correlations as reported in this subsection will be taken into account. Hence, if effects found can be attributed to scene changes, this will be noted.

4.3.2 The film fragments

We will now describe the results gathered through 24 Repeated Measures ANOVAs, with film fragments (8 levels) and time (12 time windows) as within subject factors. The results

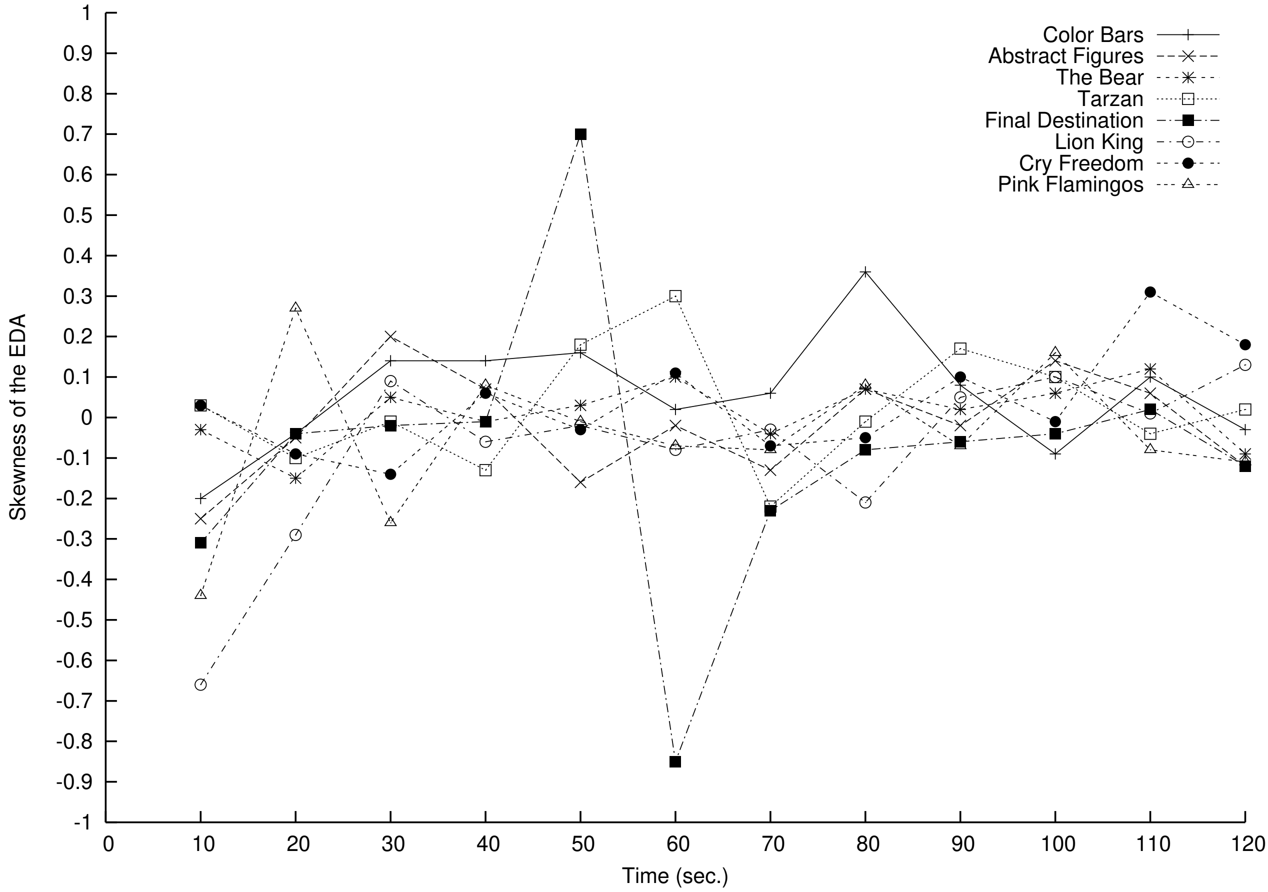


Figure 4.1: The skewness measure of the galvanic skin response / ElectroDermal Activity EDA) for each of the eight film clips.

will be presented subsequently for each of the four biosignals separately. Figures 4.2–4.5 show the mean signals over time, for each of the film fragments separately, averaged per time window of 10 seconds. As measure of effect size partial eta squared (η^2) is reported, which indicates the proportion of variance accounted for (i.e., a generalization of r/r^2 and R/R^2 in correlation/regression analysis) [211, 737].

EDA: The kurtosis ($F(7, 147) = 2.847, p = .077; \eta^2 = .119$) of the EDA was the only parameter for which a trend was found on the factor film. Both skewness ($F(11, 231) = 3.168, p = .001; \eta^2 = .131$), and kurtosis ($F(11, 231) = 2.735, p = .012; \eta^2 = .115$) indicated an effect for the factor time, and we found a trend for the standard deviation ($F(11, 231) = 2.509, p = .065; \eta^2 = .107$). An interaction effect on film*time was found for the parameters mean ($F(77, 1617) = 2.506, p = .032; \eta^2 = .107$), skewness ($F(77, 1617) = 2.015, p < .001; \eta^2 = .088$), and kurtosis ($F(77, 1617) = 1.746, p = .007; \eta^2 = .077$). By way of example, the EDA skewness data for all eight films are plotted in Figure 4.1.

EMG frontalis: No indications for differences between the signals were found between the film fragments on any of the six statistical parameters. For the factor time, the skewness of the signal of the EMG frontalis showed a clear trend ($F(11, 231) = 2.173, p = .051; \eta^2 =$

.094). However, this effect was possibly influenced by the number of scene changes, with a significant effect ($r_s = .619, p = .032$) and two trends on scene changes ($r_s = .512, p = .089; r_s = -.538, p = .071$). No interaction effects for film*time were found.

EMG corrugator supercilii: The kurtosis ($F(7,147) = 5.793, p = .002; \eta^2 = .216$) of the signal indicates that the choice of film is a factor of influence. The factor time and the interaction of the factors film*time did not reveal factors of influence.

EMG zygomaticus major: All of its six statistical parameters discriminated between the eight film fragments: mean ($F(7,140) = 6.968, p = .001; \eta^2 = .258$), absolute deviation ($F(7,140) = 6.556, p = .001; \eta^2 = .247$), standard deviation ($F(7,140) = 5.545, p = .004; \eta^2 = .217$), variance ($F(7,140) = 2.998, p = .062; \eta^2 = .130$), skewness ($F(7,140) = 6.266, p < .001; \eta^2 = .239$), and kurtosis ($F(7,140) = 3.114, p = .022; \eta^2 = .135$). For the factor time, the skewness of the signal of the EMG zygomaticus major showed a clear trend ($F(11,220) = 2.049, p = .052; \eta^2 = .093$). In addition, for the parameters mean ($F(77,1540) = 3.148, p = .001; \eta^2 = .136$), absolute deviation ($F(77,1540) = 2.566, p = .012; \eta^2 = .114$), and standard deviation ($F(77,1540) = 2.276, p = .022; \eta^2 = .102$), an interaction effect film*time was present. It should be noted that these effects were possibly influenced by scene changes; see also Section 4.3.1.

The analysis reported in this subsection revealed various indicators for differences between the eight film fragments. In particular, skewness and kurtosis discriminated for all four signals recorded, even when taking into account possible influences of scene changes. In addition, we found several significant main effects of time, as well as significant interaction effects between time and film, underlining the variety of ways in which emotions evolve over time in the film fragments. However, this does not denote which events or features of the film fragments caused these differences. In order to achieve that, the behavior of the four signals was analyzed over time and mapped upon the full transcription of the film fragments. These analyses are described in the next section.

4.3.3 Mapping events on signals

In addition to the previous statistical analysis, we present analyses that relates the transcripts of the film fragments to the signals behavior, following the principle of triangulation; that is, "*the strategy of using multiple operationalizations or constructs to help separate the construct under consideration from other irrelevancies in the operationalization. At its simplest level, triangulation refers to the use of multiple measures to capture a construct. The triangulation strategy, however, also can be applied to multiple operationalizations of treatments and manipulations and to the use of multiple theories, analyses, analysts, methodologies, and research designs, to name but a few.*" [273]. Adopting this research strategy, we aim to generate a rich interpretation of the biosignals in terms of emotions.

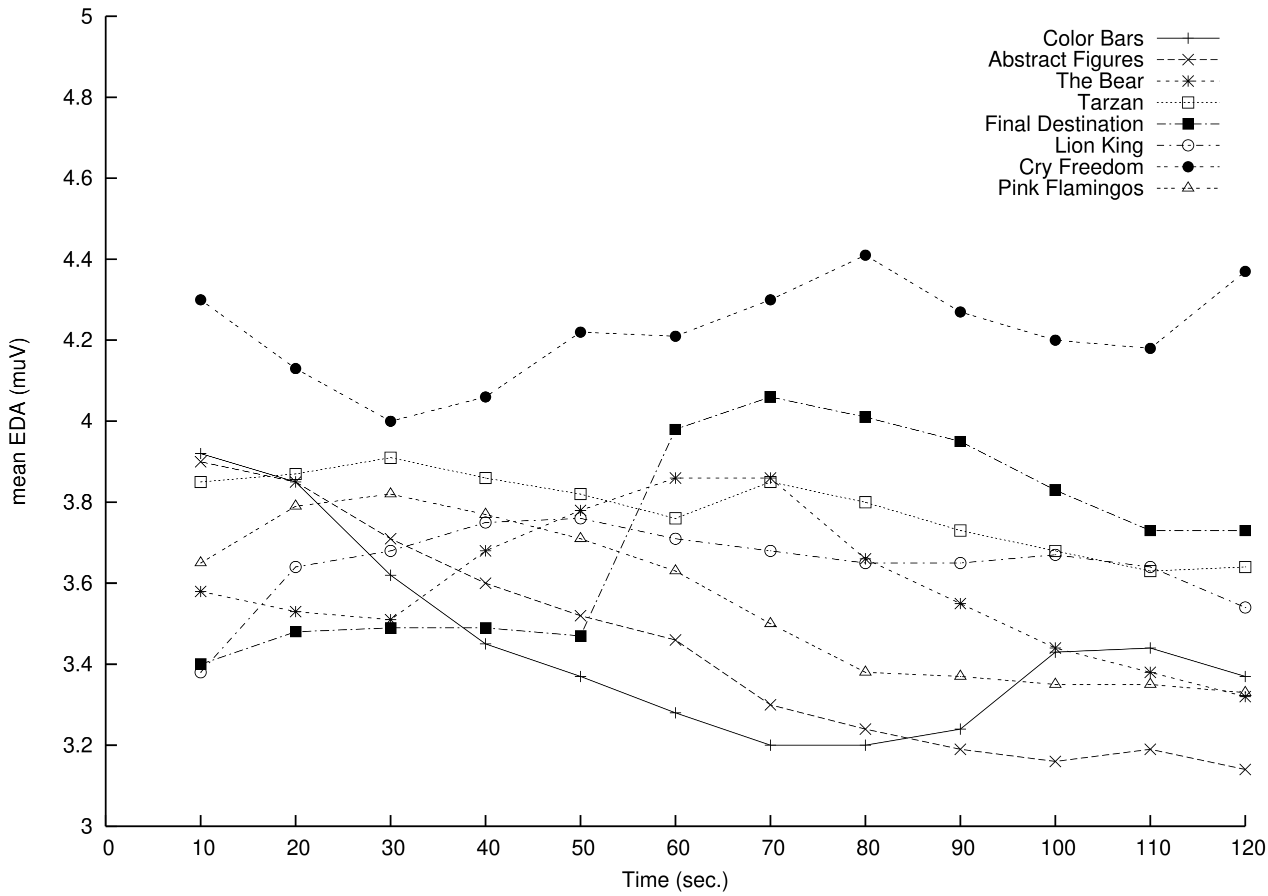


Figure 4.2: The behavior of the mean EDA signal over time, for each of the eight film fragments.

For each of the film fragments separately, each of the signals were mapped to the content of the film fragment. We will only denote substantial changes in the signals (i.e., two or more mean absolute errors from the mean of the fragment) and events of importance in the story line or the editing characteristics of the film fragments and specify these. The physiological measures are presented in four separate figures, averaged per time window of 10 seconds, respectively: the mean EDA (Figure 4.2), usually associated with arousal [62, 203, 437, 497, 530, 536, 577], the EMG frontalis (Figure 4.3) denoting fatigue [133, Chapter 10], [291], the EMG corrugator supercilii (Figure 4.4) indicating negative emotions [380], and the EMG zygomaticus major (Figure 4.5) for positive emotions [380].

Color Bars: Typical events or screenshots were absent. The EDA showed a gradual decline, indicating a decline in arousal, as is shown in Figure 4.2. The EMG signals were all stable, as shown in Figures 4.3–4.5. The skewness of the EDA (see Figure 4.1), EMG corrugator supercilii, and EMG zygomaticus major each showed one peak.

Abstract Figures: A gradual decline of the mean EDA was observed. The signal of the EMG frontalis was stable; see Figure 4.3. The signal of the EMG corrugator supercilii showed a slow increase (See Figure 4.4) and the signal of the EMG zygomaticus major was

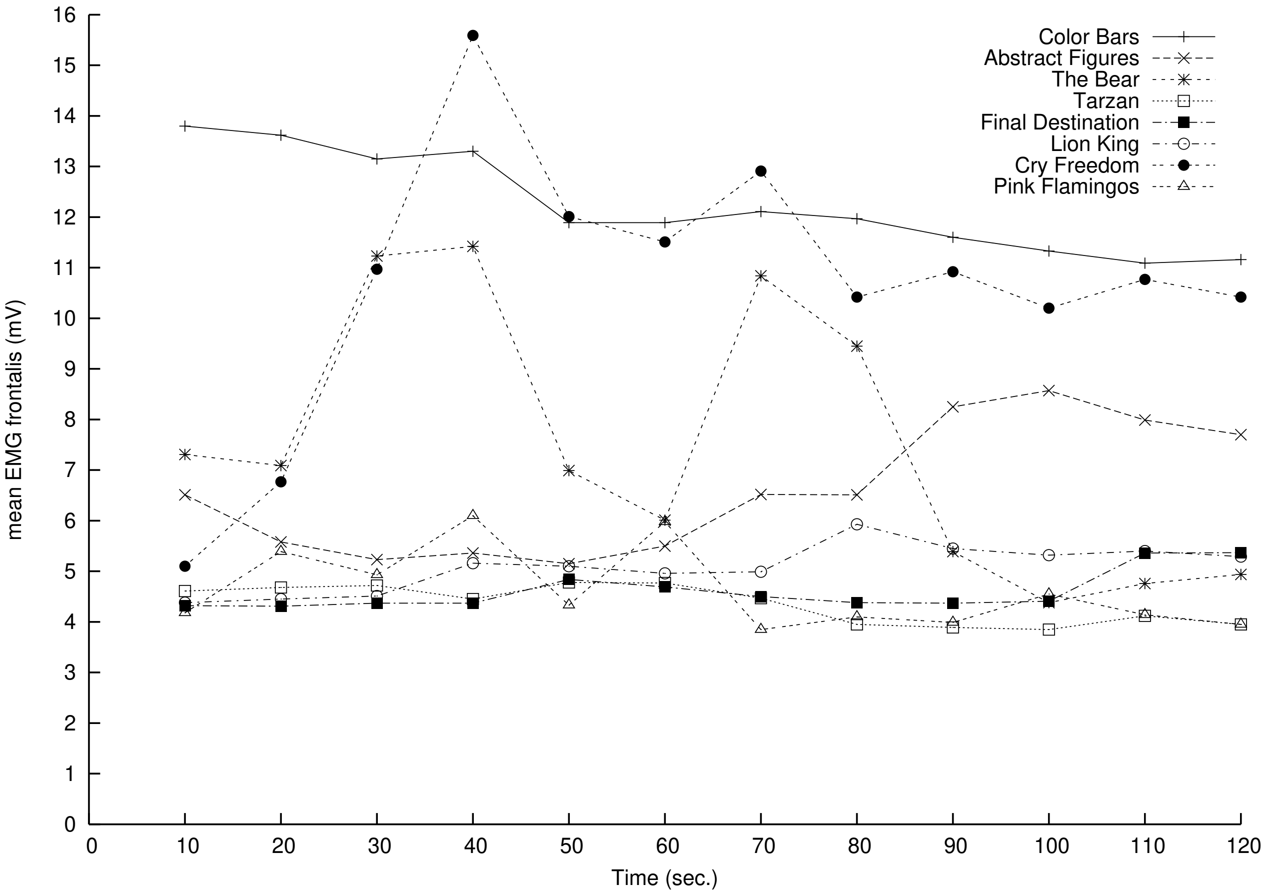


Figure 4.3: The behavior of the mean electromyography (EMG) signal of the frontalis over time, for each of the eight film fragments.

stable (see Figure 4.5). Altogether, the signals recorded for both films categorized as neutral (i.e., Color Bars and Abstract Figures) showed a similar behavior. Also for the Abstract Figures, the signals indicate a decline in arousal (through the EDA) accompanied by little fatigue, as indicated through the EMG signal of the frontalis, as shown in Figure 4.3. The skewness of the EDA and EMG frontalis each showed one peak. The skewness of the EMG corrugator supercilii showed two peaks.

The Bear: The subjects' arousal, as indicated by the EDA, increased (see Figure 4.2) and was accompanied by a frown, as measured through the EMG corrugator supercilii, up to the moment that the bees appear to be a positive signal instead of a negative one; see Figure 4.4, time: 70. Throughout the fragment, a constant varying mental workload was present, as illustrated through the signal of the EMG frontalis; see Figure 4.3. The frontalis'signal showed a peak on 50 seconds, accompanying the shot of the bear's foot. The variability of the EMG zygomaticus major (see Figure 4.5) in general indicates various moments of positive emotions. The skewness of both the EMG frontalis and the EMG corrugator supercilii showed a peak between 50 and 60 seconds denoting fatigue and negative emotions. This can be explained by the close-up of the big bear, while he was scratching. Between the 70th and 80th

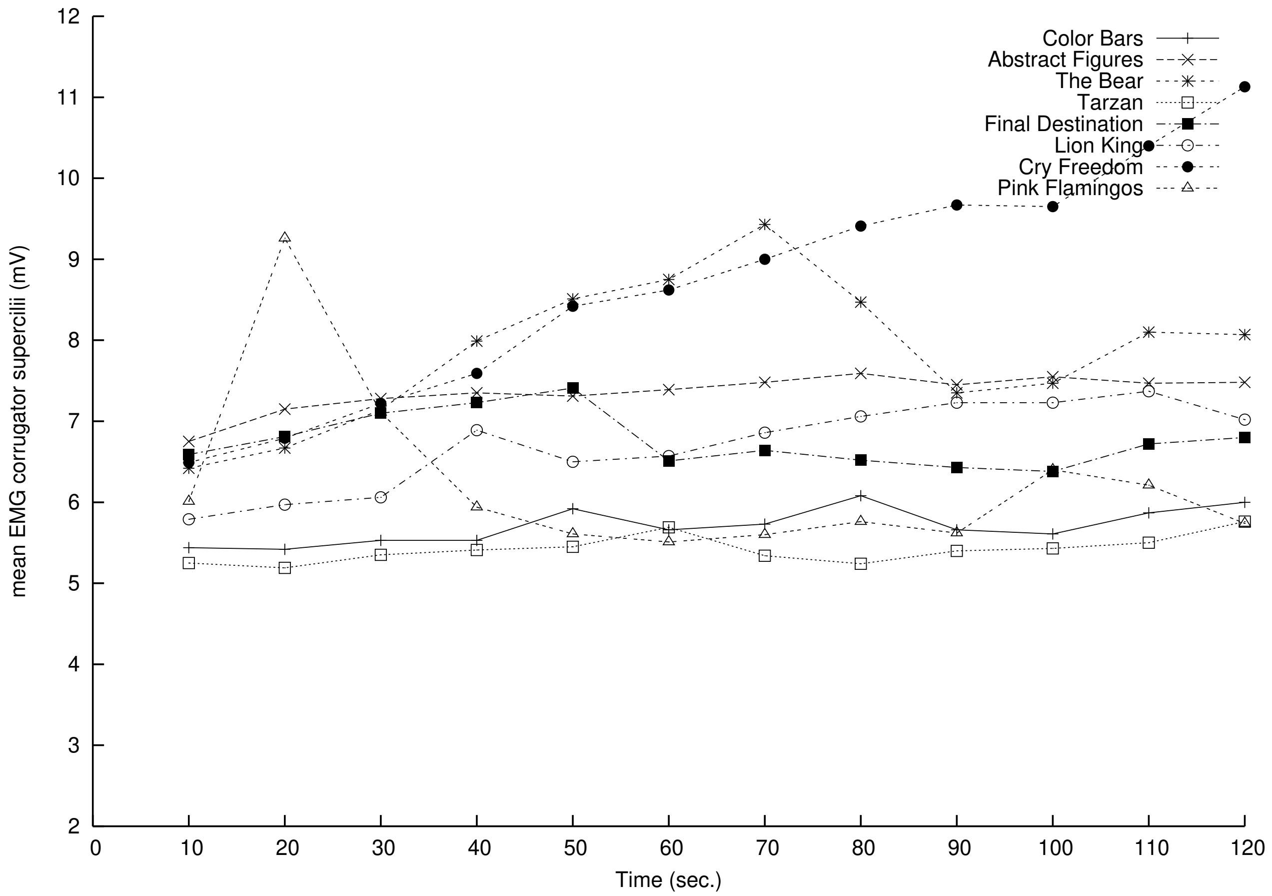


Figure 4.4: The behavior of the mean electromyography (EMG) signal of the corrugator supercilii over time, for each of the eight film fragments.

seconds participants smiled (skewness of the EMG zygomaticus major) because the mother bear and her child cub were playiing.

Tarzan: The high level of activity of the zygomaticus major (see the peak at $20-30$ seconds in Figure 4.5), indicates the presence of positive emotions. Moreover, this scene did not yield to high workload, which is not that strange for a film made for children. Participants smiled when the music was started as denoted by a peak ($10-30$ seconds) in the skewness of the zygomaticus major. The monkeys start drumming using kitchen material, which caused a peak in the skewness of both the EMG frontalis and EMG corrugator supercilii, denoting respectively fatigue or influence of scene changes and a decline in negative emotions.

Final destination: The fragment that was used has a length of 52 seconds. During the remaining 68 seconds a blue screen was shown; see also Section 3.5.2. From the start until the end of the fragment (and the start of the blue screen), a constant EDA and its skewness was determined (see Figures 4.1 and 4.2). At the end of the film fragment, a bus drives over a lady. This last event illustrated by the increase in arousal, measured through the EDA and its skewness (see Figures 4.1 and 4.2), and the laughter around the event and the immediate disappearance of the smile after the event and the end of the film fragment (time:

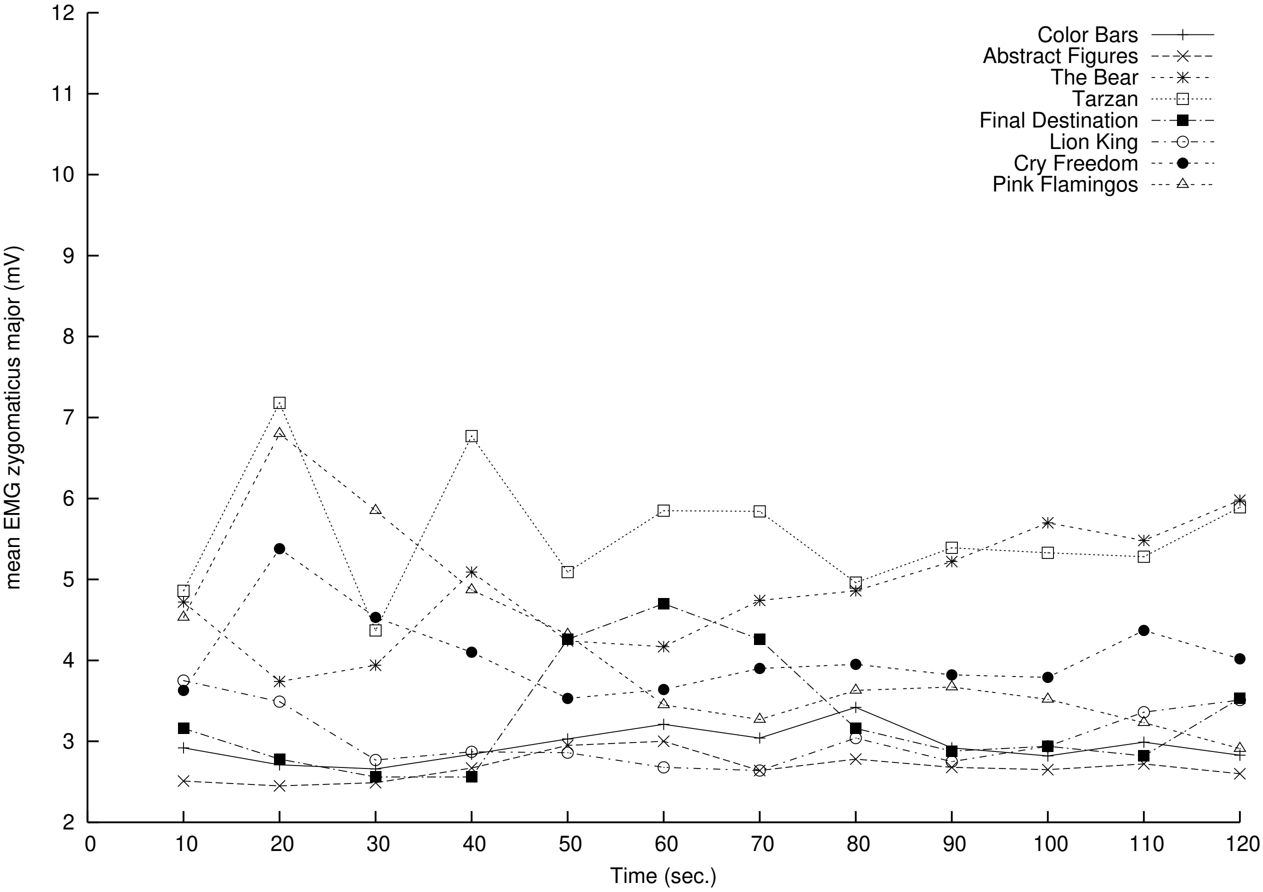


Figure 4.5: The behavior of the mean electromyography (EMG) signal of the zygomaticus major over time, for each of the eight film fragments.

50 seconds), as recorded through the zygomaticus major, presented in Figure 4.5. The effects on the EMG corrugator supercilii, however, can also be influenced by scene changes; see also Section 4.3.1. The almost simultaneous activation of the EMG corrugator supercilii and the zygomaticus major underline the presence of a mixed emotion, as was expected. The mental workload was stable over time as illustrated through the EMG frontalis signal, as shown in Figure 4.3. All 4 signals show a significant change in skewness between $40 - 50$ seconds. In this time window, a turbulent and strange situation is present in which the tension rises, as is illustrated (among other things) by the statement: *"drop dead!"* Moreover, the change of all signals in parallel denotes the hypothesized mixed emotions. After this the fragment ended and the blue screen started; no changes in the signals were recorded anymore.

Lion King: The fragment chosen from this film is a sad one: Simba (the young lion) finds his father dead. During this scene, all mean biosignals were stable, reflecting no change in emotions. However, the skewness of the EDA (see Figure 4.1) and the EMG corrugator supercilii and EMG zygomaticus major all peaked denoting fatigue, accompanied by mixed feelings, as was hypothesized. In addition, at time 60 seconds, a peak was present in the skewness of the zygomaticus major signal, illustrating the appealing shots in which the young Lion King is anxious for his father. Between the 80th and 90th seconds, a new shot is shown providing an overview, which needs to be processed, illustrated by the peak in the skewness of the EMG frontalis. Also for this fragment, aimed to trigger mixed emotions, the skewness of the signals changed in parallel, denoting such mixed emotions.

Cry Freedom: A constant tension and a large number of scene shifts is present in this film's fragment. Both were expected to contribute to the constant high arousal, as reflected in the EDA signal (see Figure 4.2). An increase in mental workload with a peak at 40 seconds (as registered by the EMG frontalis; see Figure 4.3) was present when the soldiers appeared. Next, it became clear what was going to happen, this was accompanied with a decline of the EMG frontalis signal (see Figure 4.3) indicating a decrease in fatigue. The tension / unpleasant feeling was present throughout the complete film fragments, as was illustrated by the constant increase of the EMG signal of the corrugator supercilii, as shown in Figure 4.4. However, the EMG corrugator supercilii has also been influenced by the fragment's scene changes; see Section 4.3.1. The fragment did not appeal to positive emotions; consequently, the EMG of the zygomaticus was stable (see Figure 4.5). The skewness of EMG frontalis and the EMG corrugator supercilii changed in time window $10 - 30$ seconds. This illustrates the impact of the shots chosen by the director: the boy and the crowd in an atmosphere of severe tension. Between $90 - 110$ seconds the first shots are fired and the crowd starts running, which caused arousal, negative emotions, and participants' smile fading.

Pink Flamingos: The fragment taken from this film can be best depicted as absurd. A constant arousal was present, which declined after the scene's end (at 30 seconds; see Figure 4.2). The EMG frontalis had a stable signal throughout the scene (see Figure 4.2), indicating a constant fatigue, except for a peak at 20 seconds. The absurd character of the sequence of strange shots (a huge drag queen and her tiny dog) followed by a shot of the drag queen eating inedible repulsive substances ($0 - 20$ seconds) triggered both negative and positive emotions. Consequently, the signal of both the EMG corrugator supercilii and zygomaticus major peaked as the initial smile disappeared (see Figures 4.4 and 4.5). Between $10 - 20$ seconds a high skewness of the EDA and the EMG frontalis was present, denoting high arousal and heavy processing, which nicely maps on the rapid changes in shots and the absurd events that happen. Note that most parameters of the 3 EMG signals were influenced by the rapid scene changes in the fragment, as is denoted in Section 4.3.1.

4.4 Discussion and conclusion

4.4.1 Interpreting the signals measured

In line with the previous chapter, the parameters skewness and kurtosis proved to be strong discriminating features of the EDA and EMG signals, also in the short-term analyses, which is well illustrated by the various significant effects found. They revealed compelling evidence for the distinct character of the affective signals over time for each of the film fragments. The kurtosis of the EDA and EMG zygomaticus major signal differentiated between the 8 film fragments. The skewness of all 4 signals indicated an influence of the factor time. In addition, various other statistical parameters indicated differences between the film fragments. Even with all of its statistical parameters the EMG zygomaticus major differentiated between the film fragments. Apparently, differences in emotions or feelings are usually reflected in various statistical parameters, but not necessarily in all of the ones tested (cf. [81, 82, 278]). Why the effect over all statistical parameters only occurred for the zygomaticus major activity in the present experiment is not a priori clear; maybe effects of signal resolution played a role, maybe it is related to the fact that positive emotions are more overtly expressed in our culture [450]. This last explanation could be tested by measuring the orbicularis (i.e., the "smiling muscle") in similar situations [82, 84, 85, 179].

In contrast with our findings, Larsen, Norris and Cacioppo [380] concluded that valence influenced the corrugator supercilii more than the zygomaticus major when participants experienced standardized affective pictures, sounds, and words. This can be explained by two major differences between both studies: different statistical parameters were tested and different stimuli were used; dynamic, multimodal film fragments vs. affective words, sound, or pictures [570, 663]. This issue is a matter of the traditional trade-off between, on the one hand, ecological validity, as is required for consumer applications and, on the other hand, a high level of control, which enables the isolation of effects through various factors; see also [63, 235, 237, 570, 700, 701].

The events that occurred throughout the scenes chosen from the eight film fragments were clearly reflected in the four biosignals, although not always simultaneously - as expected. Moreover, the nature of the events and the emotions they are expected to trigger, explain the clear distinction between the film fragments, as found in the analyses. Even for the two film segments that lasted considerably shorter than 120 seconds, we saw mainly decaying signals after the actual film fragment has stopped. The only exception is a zygomaticus activity at 60 seconds in the Pink Flamingo fragment, which might well be due to hilarious retrospective consideration of this absurd film clip (see Section 4.3.3 for a description). In general, however, the analyses sustain the relations between cognitive and emotional constructs and the four biosignals used, as are known from the literature. The EDA signal indicates the extent of arousal that is experienced by subjects [62, 203, 437, 497, 530, 536, 577]. However, the cause underlying this arousal can be the experienced emotions as well as the information, in terms of sensory input that has to be processed; for example, sounds and changes in scenes. The relation between the EMG frontalis and mental workload [291, 650], fatigue, or relaxation was not clearly expressed in the results obtained from this research. The EMG signal of the corrugator supercilii correlates with frowns, as is reported in the literature [380]. However, a frown can be expressed for various reasons, among which the coarse collection of negatively experienced emotions. The relation between the behavior of the EMG signal of the zygomaticus major and positive emotions is most transparent of all four signals [23, 380]. Although, the notion of positive emotions is abstract and its position within the emotion models is subject to debate [396, Chapter 29], [205, 423, 598]. Moreover, recent research revealed a relation between affective valence and working memory [230]. Hence, possibly even the EMG corrugator supercilii and the EMG zygomaticus major are influenced by information processing in addition to our emotions [171, 250, 291, 601].

With complex stimuli as film fragments, mapping events on biosignals is of importance. Moreover, it emphasizes the importance of timing for processing emotions. A delay in processing emotions larger than 10 seconds would result in strange interpretations and, consequently, situations, as our analyses show. Whether a delay of 10 second is sufficient, cannot be derived from this data however, but needs to be researched in other contexts, especially ones that are more interactive. Moreover, timeliness is especially important for the interpretation of mixed emotions: the initiation of multiple emotions [79, 92, 379]. Although mixed emotions have been a topic of debate for more than a decade [79], no accurate definition of mixed emotions exists. Do multiple emotions co-exist in parallel, is their appearance in sequence, or do they merge into each other? Two of our film fragments are suited to explore this issue, Final Destination and Lion King , since they were judged to trigger mixed emotions by the subjects. During the Lion King, the mean physiological measures did not indicate changes in emotions, although the questionnaires indicated otherwise with the questions on negative and positive emotions. The mapping of the Final Destination's transcription and the biosignals illustrated the appearance of mixed emotions: an increase in arousal and frowning was shown, immediately followed by a smile. The onset of these emotions clearly differed, with the corrugator activity linked to negative feelings mostly preceding the zygomaticus activity linked to positive feelings. However, the skewness measure of the signals seems to indicate the existence of mixed emotions perfectly with both film fragments. The skewness of the EDA, EMG corrugator supercilii, and the EMG zygomaticus major in parallel showed a clear reaction to both films on one event, indicating a change in arousal, negative, and positive emotions. This would argue for the parallel occurrence of emotions. However, follow-up research is needed to verify whether or not the current findings will sustain throughout various groups of participants, stimuli, and settings. In particular, analyses using short time windows and annotated dynamic multimodal stimuli, such as used in the research presented in this chapter, should be utilized for this purpose. This would enable ecologically valid controlled research on *ASP*.

4.4.2 Looking back and forth

The main intention of this chapter was to investigate to what extent emotional effects can be followed over time with baseline-free analysis, omitting corrections for personality traits and anomalies of the data. The figures and their interpretation clearly show - at least for the predominantly female participants in our study - that such effects are reflected in the recordings. Two points add to this expectation: First, the data we presented were averages over participants, using the raw biosignals, and possible differences between the individuals and their individual experiences were not even explored in our analyses. Thus, individual reactions might be even larger. Second, in real life the physiological reactions to the most relevant emotions might well be more marked than those elicited by excerpts from movies, as in the present experiment. Both considerations feed the hope that in relevant real-life situations, the physiological reactions of individuals can be detected through baseline-free analysis. The topic of inter-individual differences in timing patterns is also interesting in itself: one possible way to investigate it would be an experiment in which subjects are repeatedly presented with the same sequence, thus reducing noise.

All in all, this chapter underlines that emotion-aware consumer products could become a reality, as far as *ASP* and consumer electronics are concerned. For any given application, it will be necessary to investigate what processing delays are still acceptable. Additionally, the success of such an application will depend on ethical issues and issues of trust: will people feel comfortable knowing that their emotions are known, if not to others, then at least to the product? Or will the products make them feel observed and uneasy? [13, 200, 280, 525] This will no doubt be related to the context of use: at home people are usually more comfortable in having and showing emotions than at work.

In this and the previous chapter, we explored the feasibility of emotion aware consumer products that use biosignals. Consumer products require reliable (i.e., unobtrusive, robust, and lacking calibration) short-term emotion assessment. To explore the feasibility of this, an experiment was conducted where the EDA and 3 EMG signals (i.e., frontalis, corrugator supercilii, and zygomaticus major) were measured. The unfiltered biosignals were processed and six statistical parameters were derived. For each of the four biosignals, both skewness and kurtosis discriminated between affective dimensions, accompanied by four other parameters (i.e., mean, mean absolute deviation, standard deviation, and variance) that depend on the signal. The skewness parameter was also shown to indicate mixed emotions. Moreover, a mapping of events in the fragments on the signals showed the importance of short-term emotion assessment in addition to emotion assessment over several minutes as was applied in the previous chapter.

In the next two chapters, in Part III, we use another biosignal (i.e., ECG) and combine it with an indirect biosignal (i.e., speech). Although such a combination may sound obvious, it is rarely applied (see, Chapters 1 and 2). The studies conducted in Chapter 5 and Chapter 6 are identical, except for their stimuli. In Chapter 5 a subset of the IAPS database is used, which is considered as a, or even *the*, standard in *affective computing* research. Subsequently, in Chapter 6, the movie fragments used in Chapters 3 and 4 are used again.

III. Bi-modal ASP

5

Emotion models, environment, personality, and demographics

Abstract

Emotions are a crucial element for personal and ubiquitous computing. What signals to sense and how to sense them, however, remains a challenge. This study explores the rare combination of speech, electrocardiogram, and a revised Self-Assessment Mannequin (SAM) to assess people's emotions. 40 People watched 30 International Affective Picture System (IAPS) pictures in either an office or a living room environment. Additionally, their personality traits neuroticism and extroversion and demographic information (i.e., gender, nationality, and level of education) were recorded. The resulting data was analyzed using both basic emotion categories and the valence-arousal model, which enabled a comparison between both representations. The combination of heart rate variability and three speech measures (i.e., variability of the fundamental frequency (F0), intensity, and energy) explained 90% ($p < .001$) of the participants' experienced valence-arousal, with 88% for valence and 99% for arousal (ps $< .001$). The six basic emotions could also be discriminated ($p < .001$), although the explained variance was much lower: 18%-20%. Environment (or context), the personality trait neuroticism, and gender proved to be useful when a nuanced assessment of people's emotions was needed. Taken together, this study provides a significant leap toward robust generic ubiquitous *affective computing*.

This chapter is based on:

Broek, E.L. van den (2011). Ubiquitous emotion-aware computing. *Personal and Ubiquitous Computing, 15(). [in press]*

5.1 Introduction

It has been 40 years since Skinner [614] said: *The application of the physical and biological sciences alone will not solve our problems because the solutions lie in another field. . . . It is not enough to "use technology with a deeper understanding of human issues," or to "dedicate technology to man's spiritual needs," or to "encourage technologists to look at human problems." . . . What we need is a technology of behavior. . . . But a behavioral technology comparable in power and precision to physical and biological technology is lacking . . .* (p. 4-5).

Since Skinner's words [614], much has changed but even more has not. On the one hand, *phenomena of private experience, whether they be characterized as mental or emotional, conscious or unconscious, are inaccessible to direct public observation; the actions of living organisms, on the other hand, can be observed directly and studied in relation to antecedent conditions in the same way as the phenomena treated in other sciences* (p. 3) [482]. This was the case four decades ago and still is the case, despite the impressive progress of cognitive sciences and neuroscience [569] (e.g., brain imaging techniques [419, 718] and brain-computer interfaces [93, 146, 637]). On the other hand, technologies ranging from biosensors to robots have become smaller, even miniaturized [479], and can be integrated into virtually all products (e.g., clothes [622] or our homes [706]). Consequently, new branches of science and engineering have emerged, such as personal and ubiquitous computing (UbiComp) [207, 363], ambient intelligence (AmI) [120], pervasive computing [59]), wearable computing [10], and the Internet of Things [224, 358] (also known as physical computing, haptic computing, and things that think).

The true potential of the emerging branches of science such as UbiComp and AmI is more than an engineering paradigm. The envisioned systems can only be realized if human behavior can also be analyzed automatically. Subsequently, the resulting knowledge can be utilized for the integration of humans' inputs and outputs with those of their media. This yields *intuitive computing* and brings us to one of its core notions: (human) emotions. This has long been accepted by psychologists but has only been embraced by science and engineering since Picard's book *Affective computing* in 1997 [521]. As a result, ambient sensing of emotions [676], emotion-aware consumer products [679], and affective interaction [189] have been proposed. This is what this chapter will be about.

Emotions can be transmitted either overtly (e.g., by the face, the body, or the voice), covertly (e.g., biosignals), or in a combination of both [152]. On the one hand, the complex nature of emotion is illustrated by the absence of an integral model of (human) emotions (cf. [152]). So, it is still largely unknown what cues humans process in determining others' emotional states. On the other hand, the processing of signals related to emotions has been a topic of research for more than a century [144, 371, 396]. Until the end of the 20th century, attention to this relation, however, was limited and, hence, so was progress [152] (cf. [139]).

In this chapter, five issues will be addressed that are troubling the development of ubiquitous emotion-awareness:

1. A broad range of physiological signals, speech, and computer vision techniques are employed to determine people's state of emotions. Regrettably, despite the rapid growth of such techniques, these methods are either obtrusive, sensitive to noise, or both [191, 680].
2. What emotions are, how they can be described, and how they are expressed remains difficult to define [144, 302, 396, 482].
3. Although it is generally agreed that environment (or context) is of the utmost importance [6, 32, 325], lab and field research in this field is seldom compared [327, 384, 680].
4. Personality traits are seldom taken into account (e.g., [338, 524, 739]), although widely recognized as being important [453, 624, 680].
5. Demographic information (e.g., age [435, 553], gender [361, 718], culture [56, 239, 450, 470], social class [239, 470], and nationality [458]) and ethnics [585, Chapter 28], [56, 603] are known to possibly influence experienced emotions and their accompanying physiological responses. Nevertheless, this basic information is often disregarded.

By addressing the combination of these issues, we expect to contribute significantly to emotion-aware technology.

In the next section (Section 5.2), we will briefly introduce the construct emotion and two models of emotion that are often used in *affective computing*. Next, in Section 5.3, we will discuss signals of emotion and introduce the hybrid approach chosen in this research. Subsequently, in Section 5.4, we will introduce the study conducted. Section 5.5 will describe how the different types of signals are processed. Next, the results and their interpretation will be described in Section 5.6. We will close, in Section 5.7, with a general discussion.

5.2 Emotions

A complete bookstore could easily be filled with books and articles on emotion and related topics. Reviewing this vast amount of literature falls beyond the scope of the current chapter. Moreover, excellent handbooks (e.g., [144, 396]) and review articles (e.g [139, 302]) have already been published on this topic. So, no overview of emotion theories and their levels of description has been provided so far nor will be provided in the remaining chapters. Instead, we will now work towards a stipulative definition of emotion. This is necessary as there is *still no consensus on a definition of "emotion," and theorists and researchers use "emotion" in ways that reflect different meanings and functions* (p. 363), as Izard recently stated [302]. Moreover, one of this chapter's main aims is to compare two emotion representations. Hence, it is needed to provide some foundation on the notion of emotions. The overview presented next takes up a special section of the journal *Emotion Review* as foundation [302].

5.2.1 On defining emotions

In the search for consensus on what emotion is, Izard [302] identified six key structures of emotion, namely: $i)$ neural systems, $ii)$ response systems, $iii)$ feelings, $iv)$ expressive behavior, signalling systems, $v)$ antecedent cognitive appraisal, and $vi)$ cognitive interpretation of feelings. Moreover, seven functions of emotions were identified: $i)$ facilitates attention and direction of responses, $ii)$ motivates cognition and action and provides emotion information [484], $iii)$ alters the salience or value of an event to facilitate adaptive associations, $iv)$ contributes to emotion and behavior regulation, well-being, and the safeguarding of sensitivities and concerns, $v)$ social signaling, communication, $vi)$ provides a neural (often conscious) workspace for assembling solutions, and $vii)$ different emotions (and their structures) have different functions. Together, these aspects and functions of emotion provide a knowledge space we can work with.

Izard [302] concludes by stating that *Emotion consists of neural circuits (that are at least partially dedicated), response systems, and a feeling state/process that motivates and organizes cognition and action [484]. Emotion also provides information to the person experiencing it, and may include antecedent cognitive appraisals and ongoing cognition including an interpretation of its feeling state, expressions or social-communicative signals, and may motivate approach or avoidant behavior, exercise control/regulation of responses, and be social or relational in nature.* (p. 367) [302]. This does not provide us with a precise definition of emotion as a unitary concept. However, it does provide us something to hold on to and work with. Moreover, more than anything else, it emphasizes both the complexity of emotions and their ubiquitous nature.

5.2.2 Modeling emotion

As we outlined in the previous section, emotions are complex to untangle. However, there is general consensus on the neural systems underlying them, which are at least partly dedicated [302]. This having been said, emotion recognition remains challenging for both man and machine. For example, different emotions and different structures of each emotion have different functions [302] and neural systems are influenced by much more than solely emotions (e.g., imagine what happens to your heart rate when you start walking).

For engineering practice a workable model of emotion needs to be adopted. However, there are good arguments to state that such a model is beyond science's current reach. Nevertheless, some model needs to be chosen; otherwise, signals of emotion cannot be processed and classified and our endeavor ends prematurely. Psychology distinguishes two emotion models: i) discrete emotion categories and ii) a (2D or 3D) continuous dimensional model of emotion.

The discrete emotion categories originate from Darwin's pioneering work on basic emotions. The theory behind this model assumes that these emotion categories are hard-coded into our neural system and recognized universally [116, 181, 391] (cf. the debate on color categories as unveiled by Berlin and Kay [550] and the notion of basic level categories coined by Hoenkamp, Stegeman, and Schomaker [284]). Although still a topic of debate, most consensus exists on the six emotion categories: happiness, sadness, surprise, fear, anger, and disgust.

The (continuous) dimensional model of emotion assumes orthogonal unipolar or bipolar dimensions that together can describe the emotional state a person is in. Most often Russel's circumplex or valence-arousal (VA) model of emotions [105, 176, 202, 452, 567, 647] is adopted. This distinguishes arousal and valence (i.e., pleasure / displeasure) as two orthogonal bipolar factors that describe emotions. The dimensional VA model has frequently been extended [79, 202]; for example, to enable the incorporation of mixed emotions [92, 458, 679, 709]. These extensions often incorporate two unipolar valence dimensions: one for positive and one for negative valence, instead of one bipolar valence dimension. Such extended VA models incorporate three dimensions, instead of two. This approach was also adopted for the current research.

5.3 Ubiquitous signals of emotion

As we already mentioned in the introduction, the techniques usually employed to process signals of emotion are often either obtrusive, sensitive to noise, or both. We will now discuss each of the three signals: biosignals, computer vision, and speech and identify their pros and cons. Next, we will introduce the hybrid approach adopted in the research described in this chapter.

Features of physiological signals (or biosignals) are known to indicate emotions [85, 191, 672]; however, measurement of such signals is often experienced as obtrusive by participants. However, with the progress of wearable computing and wireless sensing technologies in the last decade this problem quickly vanishes [10, 138, 257, 414, 438, 508, 513, 728, 744]. In parallel, biosignal recording, even with a certain amount of obtrusiveness, is embraced by the general public in Western societies (e.g., real-time ElectroCardioGram (ECG) processing to guide athletes). Hence, the path towards biosignal-based *affective computing* would seem to be paved.

An alternative for biosignals are computer vision techniques. These can be employed both as a static (i.e., image) and a dynamic technique (i.e., video) [241, 717, 727, 739]. Although appealing, computer vision techniques are only usable for emotion recognition in very stable environments. Speech-based *affective computing* is probably the most exhaustively studied technique of this triplet. Its early studies included humans as classifiers, followed by advanced statistical procedures, and, subsequently, automated digital speech signal processing by computers [182, 590, 644, 725, 739]. Speech can be considered as an indirect biosignal that is very well suited to unveil the emotional state of a person.

Signals from the first group (i.e., biosignals) are rarely combined with signals from the other two groups (i.e., computer vision and speech). In contrast, biosignals themselves are frequently combined (e.g., [247, 338, 524, 675]). Also signals from the speech processing and computer vision groups are frequently combined [131, 184, 511, 739]. However, some exceptions exist. Bailenson et al. [25] combined computer vision and physiological measures. Van Drunen et al. [682] combined physiological measures with eye tracking, thinking aloud, and user-system interactions (cf. [680]). The current study combines speech and biosignals for emotion recognition. To the author's knowledge only two groups have reported on this combination: Kim et al. [336, 337, 339, 340] and the current author and colleagues [676]. A possible explanation is the lack of knowledge of the application of this combination of measures. We expected to extract features from both the speech and the ECG signal of people's experienced valence and arousal as well as on their basic emotions. Let us now briefly introduce both of these signals.

The human speech signal can be characterized by various features and their accompanying parameters. However, no consensus has thus far been reached on the features and parameters of speech that reflect the emotional state of the speaker. Most evidence exists for the variability (e.g., standard deviation; SD) of the fundamental frequency (F0), the intensity of air pressure ($\mathcal{I}$), and the energy of speech ($\mathcal{E}$) [182, 590, 644, 725, 739]. Therefore, we have selected these speech features in the current research.

The ECG is an autonomic signal that cannot be controlled easily, as is the case with electrodermal activity [85]. ECG can be measured directly from the chest. Where Blood Volume Pulse (BVP; i.e., photoplethysmography to detect blood pressure and determine Heart Rate, HR) can already be recorded rather unobtrusively for some time, it would seem it will also be possible soon for ECG [414, 513]. Previous research identified various features of ECG as indicators for both experienced valence and arousal [14, 85, 105, 481]. However, most evidence is provided for the HR variability (HRV) [304, 349]. HRV decreases with an increase in mental effort, stress, and frustration [85, 682]. Moreover, HRV can be influenced by the valence of an event, object, or action [14, 481, 538]. On the whole, HRV as can be derived from ECG is a rich source of information and has been shown to be a powerful discriminator between emotions [304, 349]; therefore, HRV was selected as the ECG's feature.

5.4 Method

5.4.1 Participants

40 volunteers (20 male, 20 female [361]; average age 27.8; standard deviation: 7.6; range: 18-49) participated. None of them had hearing impairments or any known cardiovascular problems. All had (corrected to) normal vision. The participants were ignorant of our research goals.

The participants were divided into two groups of 20 each. One group of participants was assigned to an office environment, in which they took place in an office chair. The other group of participants was assigned to a living room environment, in which they sat on a couch. At both locations, the room was silent and darkened and a screen was placed in front of the participant. Although both environments were controlled, this enabled an operationalization of the concept context (or environment) and, hence, its influence on ubiquitous *affective computing*.

After instructions, the participant signed an informed consent, and the ECG measurement belt and headset were positioned. Next, the participant read aloud a non-emotional story to a) verify by asking whether or not the participant had understood the instructions, b) to test the equipment, and c) to determine their personal baseline for both the speech and the ECG signal.

Traditionally personality traits are often assessed using questionnaires. Although convenient in most research settings, it is not convenient with most end consumer applications. The latter was the main reason to omit personality questionnaires from the research as presented in the previous two chapters. However, with the current research aims to unveil key features for *ASP* and, hence, the former is applicable. Therefore, with the current study as well as with the study presented in the next study, personality traits are taken into account.

Using a questionnaire, we recorded general demographic information of the participants: age [390, 435, 661], level of education, and nationality [603]. This information was used to control for them as possible sources of influence [680]. Next, the participants were also asked to fill in a revised, short scale of the Eysenck Personality Questionnaire (EPQ-RSS) [187]. Two binary indices were derived from the EPQ-RSS. These indicate the participants' personality traits neuroticism and extroversion, which are both known to influence the emotions experienced [126, 422, 442, 676].

5.4.2 International Affective Picture System (IAPS)

To elicit an emotional response, the participants looked at 30 pictures from the International Affective Picture System (IAPS) [374]; see Table 5.1 for their identification numbers. The

Table 5.1: The 30 IAPS pictures [374] with the average ratings given by the participants on the positive valence, negative valence, and arousal Likert scales. From the positive and negative valence ratings, three valence categories were derived: neutral, positive, and negative. Using the scores on arousal, two arousal categories were determined: low and high. Consequently, we were able to assess a discrete representation of the valence-arousal (VA) that distinguished six compounds.

IAPS identifiers	basic	VA model	
	emotions	valence	arousal
4624, 4625, 7450, 8033, 8220	joy	positive	high
2120, 3015, 6022, 6230, 6312	anger	negative	high
5000, 5020, 5030, 5800, 7900	relaxed	positive	low
2141, 2205, 2375, 9220, 9435	sadness	negative	low
2704, 5920, 7640, 8160, 8232	neutral 1	neutral	high
2214, 7000, 7041, 7484, 9070	neutral 2	neutral	low

IAPS set is based on a dimensional model of emotion [105, 374, 452]; however, as has been shown, this set also has great potential to reflect multiple emotion categories [452, 676]. Moreover, this set of pictures has been thoroughly and repeatedly validated [374, 452] and, as such, serves as a sort of ground truth for emotion research. The pictures were randomly presented on a 15.4 inch TFT screen (1280×800 pixels, 60 Hz refresh rate; video card: ATI Mobility Radeon 9700).

Each of the 30 IAPS pictures (see Table 5.1) were shown for a duration of 20 seconds, which is more than sufficient for emotion assessment [679]. After the presentation of each picture, the participants had 30 seconds to describe it, followed by a resting period of 20 seconds. During these 50 seconds, a gray screen was shown. The experiment started and finished by displaying a gray screen lasting 50 seconds.

5.4.3 Digital Rating System (DRS)

After all 30 IAPS pictures were presented and the participants had described them, the participants had been asked to judge the IAPS pictures using a Digital Rating System (DRS). The DRS displayed the IAPS pictures to aid the participant's memory, together with 11 point (range: 0 to 10) Likert scales using radio buttons; see Figure 5.1. The complete set of all 30 IAPS pictures was presented three times in separate blocks. Within the three blocks, the IAPS pictures were presented in random order. To each block, one of the three Likert scales (i.e., positive affect, negative affect, and arousal [79, 679]; see Section 5.2) was assigned in semi-random order; that is, the second block presented the arousal scale, the first and third block presented the negative and positive valence scales in balanced order. Consequently, the possible bias in judging the IAPS pictures was limited. The DRS' Likert scales were augmented with the Self-Assessment Mannequin (SAM) [67] of which three pictures were

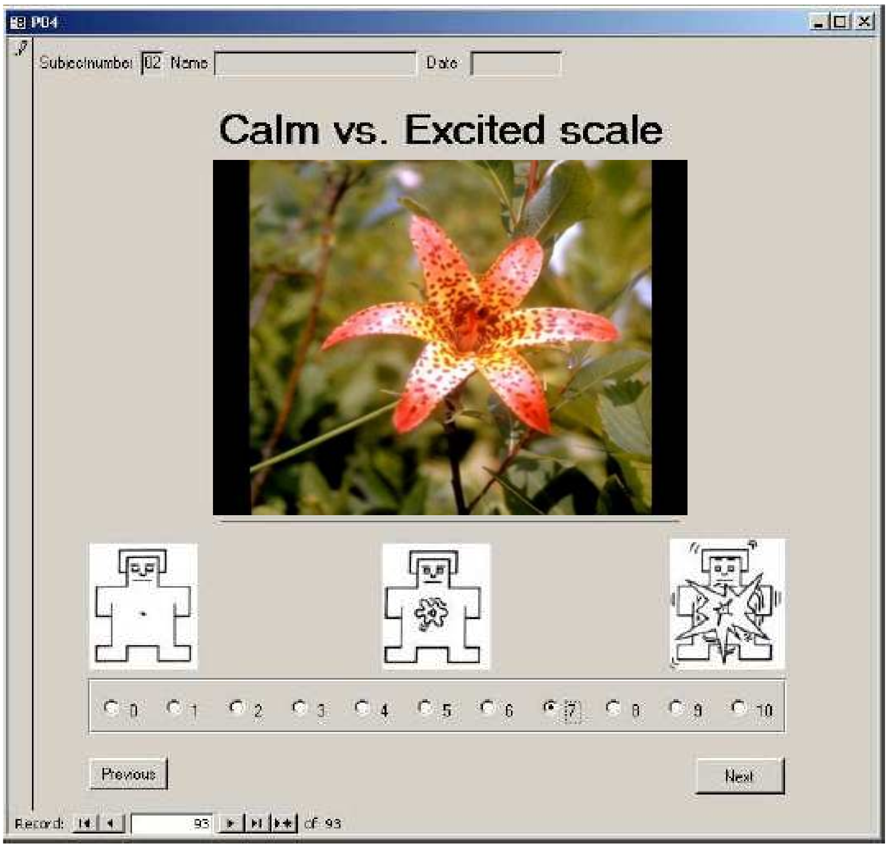


Figure 5.1: A screendump of the Digital Rating System (DRS) used in this research; see Section 5.4. An IAPS picture (category: relaxed) is shown [374]. Below the 11 point (0-10) Likert scale with radio buttons is shown augmented with three Self-Assessment Mannequin (SAM) images. With these images the experienced arousal was assessed as indicated by both the SAM images and the text "Calm vs. Excited scale".

shown; see Figure 5.1. This provided an intuitive and validated subjective assessment of the emotions the participants' had experienced.

The three scales used allowed us to contruct the VA model; see also Section 5.2. In addition, it enabled us to assign the images to the six basic emotions [105, 452], see also Table 5.1. For each picture, the average rating on each of the three scales over all participants was calculated (see Figure 5.2). This enabled a classification of the pictures into two categories (i.e., high and low) for each of the three scales: positive, negative, and arousal. From these classifications, two categories for arousal were identified: high arousal and low arousal. In addition, three categories for valence were identified: positive, negative and neutral, where the category neutral denotes neither positive nor negative valence. Table 5.1 provides a specification of the emotion categories and the IAPS images assigned to them.

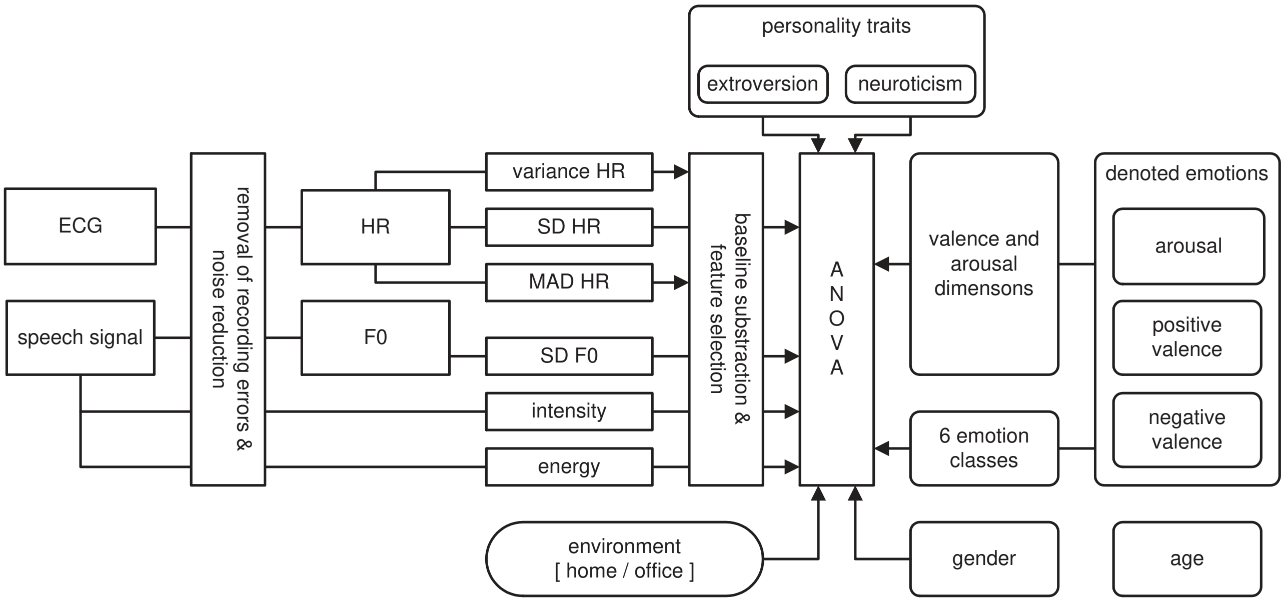


Figure 5.2: The processing scheme of Unobtrusive Sensing of Emotions (USE). It shows how the physiological signals (i.e., speech and the ECG), the emotions as denoted by people, personality traits, people's gender, and the environment are all combined in one ANOVA. Age was determined but not processed. Note that the ANOVA can also be replaced by a classifier or an agent, as a module of an AmI [694].
Explanation of the abbreviations: ECG: electrocardiogram; HR: heart rate; F0: fundamental frequency of pitch; SD: standard deviation; MAD: mean absolute deviation; and ANOVA: ANalysis Of VAriance.

5.5 Signal processing

This section describes how all of the data was recorded and, subsequently, processed (see also Figure 5.2). Speech utterances were recorded continuously by means of a standard Trust multi function headset with microphone. The recording was performed in SoundForge 4.5.278 (sample rate: 44.100 Hz; sample size: 16 bit). Parallel with the speech recording, a continuous recording of the ECG was done through a modified Polar ECG measurement belt. The Polar ECG belt was connected to a data acquisition tool (NI USB-6008). Its output was recorded in a LabVIEW 7.1 program, with a sample rate of 200 Hz.

5.5.1 Signal selection

The speech signal of three participants was not recorded due to technical problems. For one other participant, the speech signal was too noisy. These four participants were excluded from further analysis. With four other participants, either a significant amount of noise was present in their ECG or the signal was even completely absent. These participants were omitted from further processing.

Since one of the main aims was to unveil any possible added value of speech and ECG features to each other, all data was omitted from analysis of the eight participants whose ECG or speech signals were not recorded appropriately. This resulted in a total of 32 participants (i.e., 16 men and 16 women), whose signals were processed. Regrettably and surprisingly, the eight participants whose data was not processed, all participated in the office-like environment. So, 20 participants participated in this research in a home-like environment and 12 of participants sat down in an office-like environment. Conveniently, of these 32 participants, men and women were equally present in both environments.

5.5.2 Speech signal

For each participant, approximately 25 minutes of sound was recorded during the study. However, since only parts in which they spoke are of interest, the parts in which the participants did not speak were omitted from further processing. Some preprocessing of the speech signal was required before the features could actually be extracted from the signal. We started with the segmentation of the recorded speech signal in such a way that the speech signal was determined separately for each picture. Next, the abnormalities in the speech signals were removed. This resolved all technical inconveniences, such as: recorded breathing, tapping on the table, coughing, cleaning the throat, and yawning. This resulted in a 'clean' signal, as is also illustrated in Figures 5.3a and 5.3b.

After the selection of the appropriate speech signal segments and their normalization, the feature extraction was conducted. Several parameters derived from speech have been investigated in a variety of settings with respect to their use in the determination of people's emotional state. Although no general consensus exists concerning the parameters to be used, much evidence exists for the standard deviation (SD) of the fundamental frequency (F0) (SD F0), the intensity of air pressure ($\mathcal{I}$), and the energy of speech ($\mathcal{E}$) [182, 590, 644, 725, 739]. We will limit the set of features to these, as an extensive comparison of speech features falls beyond the scope of this study.

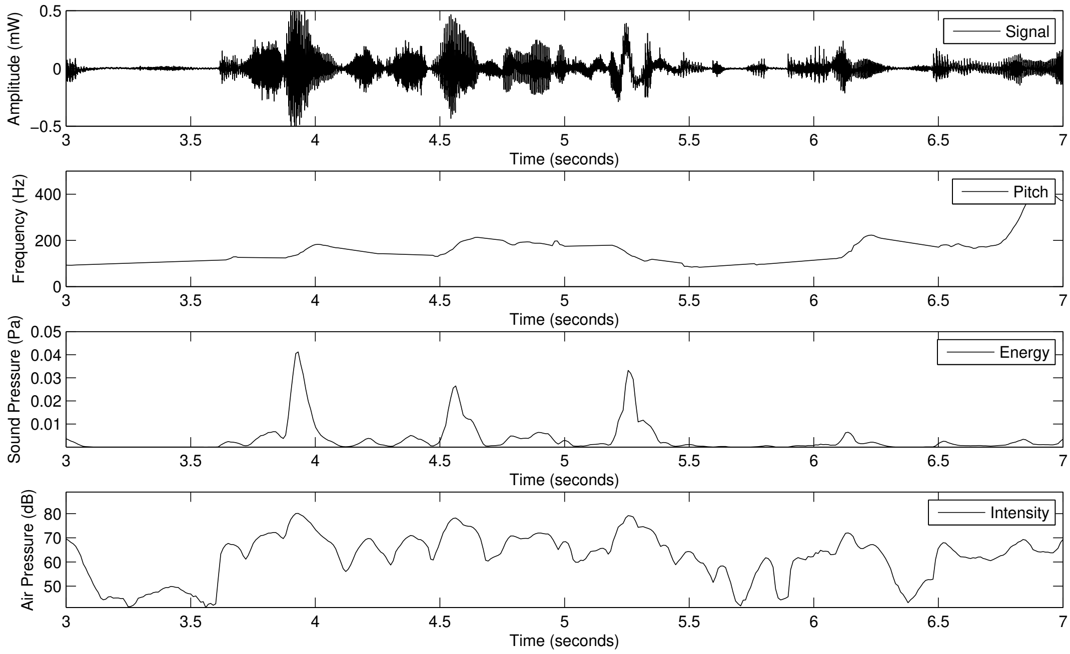
For a domain $[0, T]$, the energy $(\mathcal{E})$ is defined as:

$$\frac{1}{T} \int_0^T x^2(t)\, \mathrm{d}t, \tag{5.1}$$

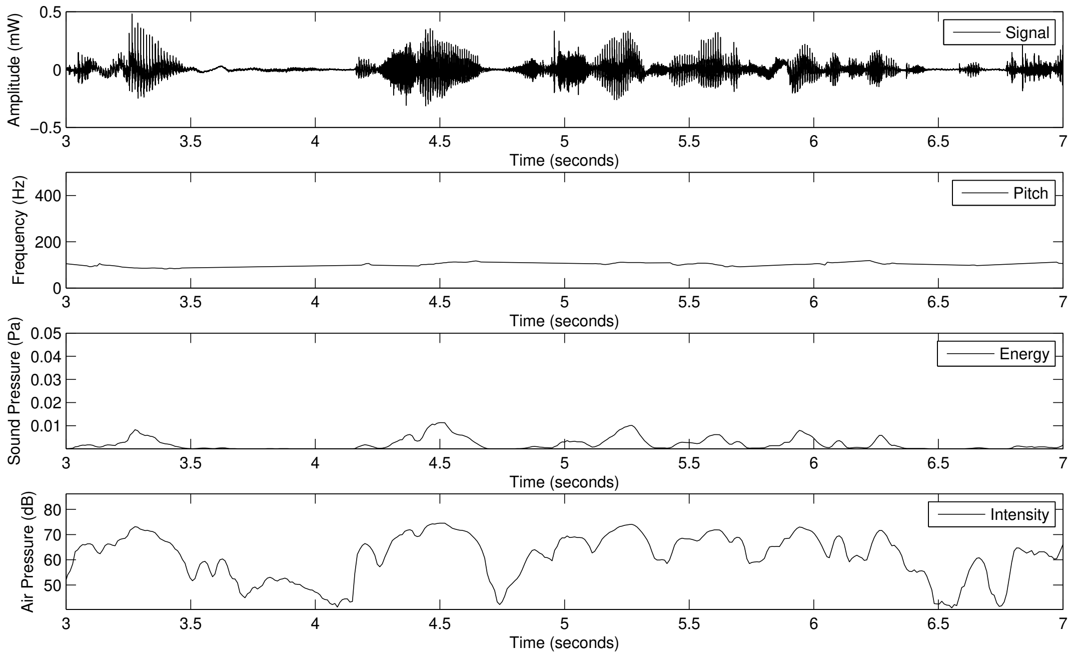
where $x(t)$ is the amplitude or sound pressure of the signal in Pa (Pascal) [54]. Its discrete equivalent is:

$$\frac{1}{N} \sum_{i=0}^{N-1} x^2(t_i), \tag{5.2}$$

where N is the number of samples.



(a) A speech signal and its features of a person in a relaxed state.



(b) A speech signal and its features of a person in a sad, tensed state.

Figure 5.3: Two samples of speech signals from the same person (an adult man) and their accompanying extracted fundamental frequencies of pitch (F0) (Hz), energy of speech (Pa), and intensity of air pressure (dB). In both cases, energy and intensity of speech show a similar behavior. The difference in variability of F0 between (a) and (b) indicates the difference in experienced emotions.

For a domain $[0, T]$, intensity ($\mathcal{I}$) is defined as:

$$10\ \log_{10} \frac{1}{T\ \mathrm{P}_0^2} \int_0^T x^2(t)\ \mathrm{d}t, \tag{5.3}$$

where $\mathrm{P}_0 = 2 \cdot 10^{-5}$ Pa is the auditory threshold [54]. $\mathcal{I}$ is computed over the discrete signal in the following manner:

$$10\ \log_{10} \frac{1}{N\ \mathrm{P}_0^2} \sum_{i=0}^{N-1} x^2(t_i). \tag{5.4}$$

It is expressed in dB (decibels) relative to P_0.

Both the $\mathcal{I}$ and the $\mathcal{E}$ are directly calculated over the clean speech signal. To determine the F0 from the clean speech signal, a fast Fourier transform has to be applied over the signal. Subsequently, its SD is calculated; see also Eq. 5.5. For a more detailed description of the processing scheme, we refer to [53].

5.5.3 Heart rate variability (HRV) extraction

From the ECG signal a large number of features can be derived that are said to relate to the emotional state of people [14, 304, 327, 349, 672, 676]. This research did, however, not aim to provide an extensive comparison of ECG features. Instead, the use of the combination of the ECG signal with the speech signal was explored. Therefore, one well-known distinctive feature of the ECG was chosen: the variance of heart rate (HRV).

The output of the ECG measurement belt has a constant (baseline) value during the pause between two heart beats. Each new heart beat is characterized by a typical slope consisting of four elements, called: P, Q, R, and S (see Figure 5.4). A heart beat is said to be

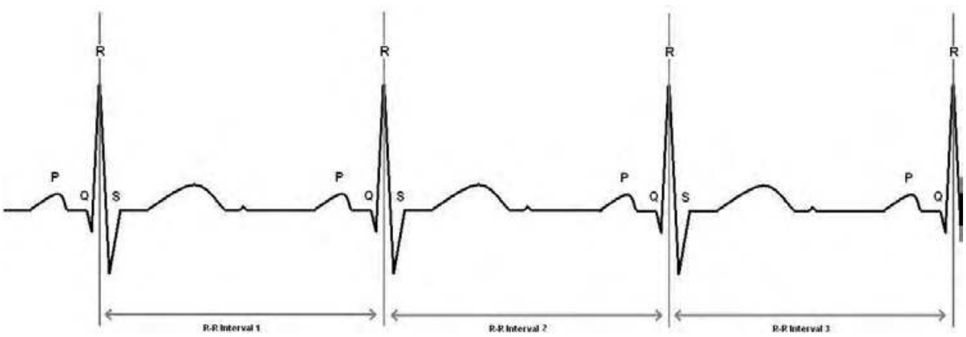


Figure 5.4: A schematic representation of an electrocardiogram (ECG) denoting four R-waves, from which three R-R intervals can be determined. Subsequently, the heart rate and its variance (denoted as standard deviation (SD), variability, or mean absolute deviation (MAD)) can be determined.

characterized by an R-wave, which is an upward deflection. The HR is calculated from the intervals between these R-waves (R-R intervals) [14, 327, 672, 676]. The measurement belt for the ECG signal appeared to be sensitive to movements of the participant. This resulted in four types of noise that can be distinguished: a heart beat that differs from the normal PQRS shape (cf. Figure 5.4), heart beats that succeed too quickly, missing heart beats in a sequence, and no HR signal at all. The ECG signal was checked automatically for all these types of noise and corrected where necessary.

The ECG signal was segmented into separate signals per stimulus, before it was processed. Next, the intervals between the R-waves (R-R intervals) of the ECG signal were determined. Subsequently, the mean R-R interval was determined. To determine the variability of the heart rate (HRV) from an ECG, the R-R intervals of the ECG were determined. Subsequently, two methods were applied for the calculation of the HRV, namely: the variance (σ^2):

$$\frac{1}{\mathcal{R}}\sum_{i=0}^{\mathcal{R}-1}(\Delta_i - \bar{\Delta})^2 \tag{5.5}$$

and the Mean Absolute Deviation (MAD):

$$\frac{1}{\mathcal{R}}\sum_{i=0}^{\mathcal{R}-1}|\Delta_i - \bar{\Delta}| \tag{5.6}$$

of the R-R intervals Δ_i. $\bar{\Delta}$ denotes the average R-R interval, and $\mathcal{R}$ denotes the number of R-R intervals. The SD of the R-R intervals is defined as the square root of Eq. 5.5: σ. Note that the σ^2 as defined in Eq. 5.5 is identical to the total spectral power. This also explains why comparisons between frequency and time domain measures have often shown that for every band of an HR power spectrum, there is at least one time-domain correlate [435]. Further, please note that various other measures were applied for the determination of the HRV [14, 304, 327, 349, 435, 672, 676]. However, with these three measures we expected to have a good indication of the use of HRV for emotion detection.

5.5.4 Normalization

To tackle intrapersonal differences in the speech signal $x(t)$, the feature series $f(t)$ calculated from this signal had to be normalized. All feature series calculated were normalized by subtracting people's personal baseline μ from the original feature series $f(t)$ [418]:

$$\tilde{f}(t) = f(t) - \mu, \tag{5.7}$$

where $\tilde{f}(t)$ denotes the normalized feature series.

The personal baseline μ was obtained at the start of the study, directly after the instructions; see also Section 5.4. This normalization is a standard correction that is often used in psychophysiological studies and has repeatedly been shown to work [85]. The resulting data is often denoted as delta or reaction score. These scores are known to be both suitable and reliable for absolute level comparisons.

The ECG signal was processed without normalization. A normalization of the features derived from the ECG signal is already included in the calculation of the HRV (cf. Eqs. 5.5 and 5.6 with 5.7). Note that for many of the other features extracted from an ECG signal, normalization is required.

5.6 Results

This section discusses the results obtained in this study. First, the considerations taken in the analysis will be presented. Next, in line with the main aims of this study, we will analyze the combined discriminative power of both the combination of features (i.e., repeated measures MANOVA (Wilks' lambda)) and the features separately (i.e., (univariate) repeated measures ANOVA (Huynh-Feldt)). This is done for two series of analysis, one to assess the dimensional VA model and one to assess six discrete basic emotions. The factors included in the analyses are: environment, the personality traits neuroticism and extroversion, and gender; see also Table 5.2.

5.6.1 Considerations with the analysis

Preliminary analysis of the ECG signal showed that the three parameters SD, the variance, and MAD of the HR (see Eqs. 5.5 and 5.6) provided similar results. This is in line with what is reported in the literature [349, 435]. Since the preliminary analysis did not reveal significant differences among the three measures for HRV and this is supported by the literature, we have chosen the most common one: the SD of the R-R intervals. Therefore, in the main

Table 5.2: Legend of the factors included in the analyses presented in Section 5.6, particular in Tables 5.3-5.6.

abbreviation	explanation
V	$v$alence
A	$a$rousal
E	$e$nvironment (or context)
P_N	$p$ersonality trait $n$euroticism
P_E	$p$ersonality trait $e$xtroversion
G	$g$ender

analysis, variance and MAD of the R-R intervals as measures for HRV were excluded; see also Figure 5.2. From this point on, the SD of the R-R intervals will simply be denoted as HRV.

The following factors were also included in the analysis: the environment (i.e., office versus living room), gender, and the two personality traits extroversion and neuroticism. Preliminary analysis showed that the other recorded demographical information (see Section 5.4) did not influence the emotional responses of the participants. Hence, this information was excluded from further analyses (cf. Figure 5.2).

To tackle the problem of finding spurious relationships that can arise when conducting multiple tests separately, first multivariate analysis was conducted and, subsequently, univariate analysis. Note that only if multivariate analysis had revealed effects was univariate analysis to be conducted to further unravel this effect. Alternatively, this problem could have been tackled by a (modified) Bonferroni adjustment, which ensures the probability of Type I errors $\leq .050$ [499]. However, this has the drawback that there is no consensus on the modification of Bonferroni.

All tests will be reported with their degrees of freedom, power, and level of significance. If the level of significance is close to zero, this will be denoted with $p < .001$, instead of providing an exact statistic. As measure of effect size partial eta squared (η^2) will be reported to indicate the proportion of variance accounted for (i.e., a generalization of r/r^2 and R/R^2 in correlation/regression analysis) [737]. The threshold for reporting results is sharp (i.e., $p \leq .010$) and, hence, ensures reliable results. Where interactions appeared that exceed the order four, they have been ignored, as they are nearly impossible to interpret. Otherwise, all interaction effects will be reported.

5.6.2 The (dimensional) valence-arousal (VA) model

In Table 5.3 the results of a repeated measures MANOVA are presented. These results relate the four features derived from the speech and ECG model onto the dimensional VA model. Results on both the integral VA model and its two dimensions are presented. Table 5.3 denotes that with the MANOVA 90% of the variance of the VA model can be explained. The results on the distinct dimensions confirm this result with respectively 88% and 99% explained variance for the valence and arousal dimensions.

When the factors environment, the personality traits neuroticism and extroversion, and gender were included in the analysis, a high level of explained variance was obtained as well. However, the power of the MANOVAs and the explained variance were much lower than when these factors were ignored; see Table 5.3. Taken together, these results confirm the efficiency of the small set of features as compiled for this research and suggest that brute force processing and extraction of large numbers of features (e.g., > 1000 [590, 727]) is not

Table 5.3: Results of the repeated measures MANOVA on the valence-arousal (VA) model and its distinct dimensions. The threshold for significance was set to p $\leq .010$.

V	A	E	P_N	P_E	G	Specification of effect
•						F(8,74) = 67.835, p < .001, $\eta^2 = .880$
•		•				F(8,74) = 3.752, p = .001, $\eta^2 = .289$
•			•			F(8,74) = 4.426, p < .001, $\eta^2 = .315$
•		•	•			F(8,74) = 2.774, p = .010, $\eta^2 = .231$
•		•			•	F(8,74) = 3.419, p = .002, $\eta^2 = .270$
	•					F(4,17) = 653.941, p < .001, $\eta^2 = .994$
	•	•				F(4,17) = 9.325, p < .001, $\eta^2 = .687$
•	•					F(8,74) = 82.962, p < .001, $\eta^2 = .900$
•	•	•				F(8,74) = 4.431, p < .001, $\eta^2 = .324$
•	•		•			F(8,74) = 4.168, p < .001, $\eta^2 = .311$
•	•				•	F(8,74) = 2.895, p = .007, $\eta^2 = .238$

required for *affective computing*.

To unravel the influence of each of the four features, a repeated measures ANOVA was executed for each of them. The results of these analyses are presented in Table 5.4. These analyses provide a much more subtle image of the features included in the MANOVA. We will now first discuss the results on the three speech features (i.e., SD F0, intensity $\mathcal{I}$, and energy $\mathcal{E}$) and, subsequently, the ECG feature HRV.

SD F0 by itself had little predictive power and even with one additional factor included no strong results were found. When both environment and the personality trait neuroticism were taken into account, an effect was found for the VA model; see Table 5.4. This effect can be mainly attributed to the dimension valence for which SD F0 was sensitive, when two factors were included; see Table 5.4. For the dimension arousal, with two levels to distinguish, no effect was found.

$\mathcal{I}$ showed to have an excellent predictive power and was able to explain almost all variance with 98% for the VA model and both of its dimensions; Table 5.4. Also strong effects were found on the VA model and its dimension when the environment and the personality trait were taken into account. However, these effects are not even close to the magnitude of the effects found when these factors were not taken into account. So, the environment, personality traits, and gender seem to have little influence, if any, on the intensity of speech as indicator for valence and arousal.

The feature $\mathcal{E}$ proved to be a good indicator for the dimension arousal of the VA model; see Table 5.4. Analysis showed that additional factors were of little influence on this effect. In addition, an effect was found on the VA model when both the environment and gender were taken into account; however, this was only marginally below the threshold of reporting. So, $\mathcal{E}$ seems to be a good and robust indicator for experienced valence.

Table 5.4: Results of the repeated measures ANOVAs on the valence-arousal (VA) model and its distinct dimensions. The threshold for significance was set to $p \leq .010$.

V	A	E	P_N	P_E	G	Specification of effect
SD F0						
•		•	•			$F(2,40) = 6.136, p = .009, \eta^2 = .235$
•		•			•	$F(2,40) = 6.327, p = .008, \eta^2 = .240$
•			•		•	$F(2,40) = 8.135, p = .010, \eta^2 = .289$
•	•	•	•			$F(2,40) = 5.924, p = .010, \eta^2 = .229$
Intensity						
•						$F(2,40) = 817.149, p < .001, \eta^2 = .976$
•			•			$F(2,40) = 13.677, p < .001, \eta^2 = .406$
	•					$F(1,20) = 1095.287, p < .001, \eta^2 = .982$
•	•					$F(2,40) = 1060.802, p < .001, \eta^2 = .981$
•	•	•				$F(2,40) = 52.611, p < .001, \eta^2 = .341$
•	•		•			$F(2,40) = 63.491, p < .001, \eta^2 = .384$
Energy						
	•					$F(1,20) = 24.123, p < .001, \eta^2 = .547$
•	•	•			•	$F(2,40) = 5.254, p = .009, \eta^2 = .208$
HRV						
•		•	•			$F(2,40) = 6.872, p = .005, \eta^2 = .256$
•		•			•	$F(2,40) = 9.463, p = .001, \eta^2 = .321$
•			•		•	$F(2,40) = 6.354, p = .007, \eta^2 = .241$
	•	•			•	$F(1,20) = 8.493, p = .009, \eta^2 = .298$
	•		•		•	$F(1,20) = 8.772, p = .008, \eta^2 = .305$
•	•	•	•			$F(2,40) = 7.426, p = .004, \eta^2 = .271$
•	•	•			•	$F(2,40) = 9.736, p = .001, \eta^2 = .327$

The ECG feature HRV is frequently used as an indicator for emotions. In the perspective of the VA model, it has been reported to indicate both the experienced valence and the experienced arousal. These results are confirmed by the current study; see Table 5.4. However, for the VA model as well as for its distinct dimensions, the factors environment, the personality trait neuroticism, and gender proved to be of influence. In contrast with the speech features, the power of HRV as indicator for the VA model was only unveiled when these factors were included in the ANOVA; see Table 5.4.

5.6.3 The six basic emotions

In Table 5.5 the results of a repeated measures MANOVA are presented that mapped the four features derived from the speech and ECG model onto the six basic emotions. The MANOVA showed an effect of the four features on the six basic emotions, with and without other factors included. The effect of the factors environment, the personality trait neuroticism, and gender were limited. The personality trait extroversion was of no influence.

Table 5.5: Results of the repeated measures MANOVA on the six basic emotions. The threshold for significance was set to $p \leq .010$.

E	P_N	P_E	G	Specification of effect	
				F(20,400) =	4.330, $p < .001$, $\eta^2 = .179$
•				F(20,400) =	2.332, $p = .001$, $\eta^2 = .106$
	•			F(20,400) =	4.777, $p < .001$, $\eta^2 = .194$
•	•			F(20,400) =	4.710, $p < .001$, $\eta^2 = .191$
•			•	F(20,400) =	4.869, $p < .001$, $\eta^2 = .196$
	•		•	F(20,400) =	3.951, $p < .001$, $\eta^2 = .166$

To gain understanding in the influence of the four features, repeated measures ANOVAs were executed for each of them. The results of these analyses are presented in Table 5.6. First, we will discuss the results of the three speech features (i.e., SD F0, intensity $\mathcal{I}$, and energy $\mathcal{E}$). Second, the results of the ECG feature HRV will be discussed. In none of the analyses did the personality trait extroversion show any influence; therefore, this factor will not be mentioned further on.

SD F0 by itself showed to have little predictive power; see Table 5.6. Only when either environment and the personality trait neuroticism or environment and gender were taken into account was an effect found. $\mathcal{I}$ showed to have no predictive power at all, neither by itself nor in combination with other factors. $\mathcal{E}$ showed to have a good predictive power by itself. The four factors included in the analysis were of no influence on $\mathcal{E}$; see Table 5.6.

Of all four features, HRV showed to have the highest predictive power. However, this was only the case when two out of the three factors included were taken into account. In each of these three cases, roughly 30% of the variance in the data could be explained.

Table 5.6: Results of the repeated measures ANOVAs on the six basic emotions. The threshold for significance was set to $p \leq .010$. For the Intensity ($\mathcal{I}$) of speech no results are reported as none of them exceeded the threshold.

E	P_N	P_E	G	Specification of effect	
SD F0					
•	•			F(5,100) =	6.292, $p = .006$, $\eta^2 = .239$
•			•	F(5,100) =	6.441, $p = .005$, $\eta^2 = .244$
Energy					
				F(5,100) =	6.352, $p < .001$, $\eta^2 = .241$
HRV					
•	•			F(5,100) =	7.078, $p = .004$, $\eta^2 = .261$
•			•	F(5,100) =	9.355, $p = .001$, $\eta^2 = .319$
	•		•	F(5,100) =	6.601, $p = .006$, $\eta^2 = .248$

5.6.4 The valence-arousal (VA) model versus basic emotions

When the VA model is compared with the basic emotions model, the following ten main conclusions can be drawn:

- Both emotion representations can handle the variation in participants, even without including additional information such as the environment, personality traits, and gender; see Tables 5.3- 5.6.
- Using the VA model a very high amount of variance can be explained: 90%. This is much higher than with the basic emotions: 18% (cf. Tables 5.3 and 5.5).
- Many more effects were found with the VA model than with the basic emotions as representation for emotions (cf. Tables 5.3 and 5.5 and Tables 5.4 and 5.6).
- The SD F0 showed to have a good predictive power with both emotion representations; see Tables 5.4 and 5.6.
- The intensity of speech ($\mathcal{I}$) is by far the most informative feature for the VA model; see Table 5.4. In contrast, with the basic emotions it has no predictive power at all; see Table 5.6.
- The energy of speech ($\mathcal{E}$) was a very good predictive power for arousal and a good predictive power for the six basic emotions; see Tables 5.4 and 5.6.
- The ECG feature HRV showed to be heavily influenced by multiple factors that were included in the analysis. However, when these are taken into account, HRV can serve as a rich source of information to unveil emotions; see Tables 5.4 and 5.6.
- The personality trait extroversion had no significant influence on the participants' experience of emotions; see Tables 5.3- 5.6.
- Gender has some influence, although limited; see Tables 5.4 and 5.6. For the speech signal this could be partly explained by the normalization of the signal.
- Although approached from another angle, both emotion representations as treated in this chapter share many characteristics. This is mainly because a discrete representation of the VA model was used that can distinguish six compounds, similar to the six basic emotions.

The current study illustrates that the representation of emotions remains a topic of debate; see also Sections 5.2 and 5.3. In practice, both discrete basic emotions and dimensional models are applied [105, 176, 202, 452]. This study compared these two representations. Data of the current study suggests that the VA model is most appropriate, as the explained variance is much higher than with the basic emotions: 90% versus 18%. As Eerola and Vuoskoski [176] state, the resolution of the discrete and categorical models is poorer. Moreover, current results provide more support for the VA model than for suggest basic emotions (cf. [202]). The discrepancy in explained variance of the present analyses (see Tables 5.3 – 5.6) can be attributed to the variation between the stimuli assigned to one basic emotion.

With both models of emotion, many interaction effects have been reported in the accompanying sections, in particular in relation to HRV. However, even twice as many effects would have been reported if the threshold for significance was set to p $\leq .050$, as is most often done. More than anything else this illustrates the complexity of people and their emotional state. Moreover, more than a choice for one of both emotion models, the current data suggests that a holistic model would be most appropriate. This also explains the variation in results reported in the literature, in particular in research that goes from lab to life [680]. Despite its drawbacks, studies that aim to bridge this gap and take into account multiple factors should be encouraged. With the current study such an attempt has been made; however, as the data illustrate, many more are needed.

5.7 Discussion

This section will discuss the results presented in the previous section further and relate them to the current state-of-the-art research. As was described in the introduction (Section 5.1), this research features five key issues, namely: i) hybrid (i.e., speech + biosignals) *affective computing*, ii) modeling emotion, iii) including environment (or context), iv) taking people's personality into account, and v) unveiling the possible importance of demographic information. Each of the these key issues will be discussed in this section. Next, a brief general discussion will be provided. We will end this chapter with some conclusions.

5.7.1 The five issues under investigation

Nowadays, not only speech and computer vision but also biosignals such as ECG can be applied completely unobtrusively [10, 138, 257, 414, 438, 508, 513, 728, 744]. Speech and biosignals are par excellence suitable for personalized and ubiquitous *affective computing* technology. However, surprisingly, this combination has hardly been explored; except for the author's own work [676] (see Chapters 5 and 6), the only work the author is acquainted with that applied this combination is that of Kim et al. [336, 337, 339, 340]. Processing both signals in parallel can, however, be done conveniently, as is illustrated by this study; see also Figure 5.2. Moreover, as is shown in this chapter, the combination of speech and biosignals provides a potentially rich source of complementary information. This was confirmed by the analyses presented in the previous section; see also Tables 5.3-5.6. True bimodal (or hybrid) including biosignals and either speech or vision-based techniques should be explored more often, despite the various methodological and technical hurdles that need to be taken for its execution. Moreover, without any doubt, trimodal (i.e., biosignals, speech, and vision-based) *affective computing* would also be fruitful.

To ensure the correct assessment of the experienced emotions of people, the IAPS set was used in the current research [105, 374, 452]. Throughout the years, IAPS has become a ground truth for emotion research, as it is repeatedly well validated. Since the representation of emotions is still a topic of debate, both the dimensional VA model and the categorical basic emotions were employed, using the same set of IAPS pictures [374, 452]. This enabled a one-on-one comparison between both emotion representations [105, 176, 202, 452, 567]. Although the various representations of emotions are frequently discussed, it is rare that two (or more) models are mapped upon affective signals (cf. [676]). However, par excellence, the setup of the current research facilitated this. The results as discussed in the previous section support both models. However, more convincing effects have been found for the dimensional VA model. Although further studies should be conducted on the mapping of *affective computing* techniques upon models of emotion, the results of the current study provide a clear support for the VA model.

That context plays its role in human functioning (e.g., information processing) is generally accepted. However, how to operationalize such an abstract concept? Moreover, is context not different for us all (e.g., because it depends on our memories)? To enable a feasible operationalization of the concept context it was brought down to the concept environment. The same study was conducted in two environments. Half of the participants participated in a living room setting and half of them participated in an office setting. This enabled a comparison between both settings. Both repeated measures MANOVAs (see Tables 5.3 and 5.5) showed a (very) sharp decline in power and explained variance when environment was taken into account as a factor. This implies that including environment as a factor introduces noise instead of an additional source of information that can explain the variance in the data. However, the (univariate) ANOVAs, with both emotion representations (a separate one for each of the four features) provide another point of view. With these analyses environment did help to unveil emotions. This suggests that the combination of features chosen for this study can handle the influence of the environment (or context) excellently. This stresses the complementary characteristics of the features chosen, as was already claimed earlier on in this chapter. In parallel, it identifies the influence environments do have on physiological responses to emotions. Follow-up research should explore this intriguing finding further.

The personality traits neuroticism and extroversion, both known to influence the experience of emotions [126, 422, 442, 453, 676], were assessed to determine their relation to the affective state of the participants. Independent of the emotion representation chosen, the personality trait extroversion has shown to be of hardly any influence. This is in line with an earlier study by the author [676] but deviates from other literature [126, 422, 442]. In contrast, the personality trait neuroticism has shown to be of influence, with both emotion representations. However, its influence depended heavily on the emotion representation chosen. With the dimensional VA model, the repeated measures MANOVA (see Table 5.3) showed a (very) sharp decline in power and explained variance when environment was taken into account as a factor. As with the environment, this implies that including the personality trait neuroticism as a factor introduces noise instead of an additional source of information that can explain the variance in the data. In contrast, with the six basic emotions, the repeated measures MANOVA (see Table 5.5) showed a small increase in both power and explained variance when environment was taken into account as a factor. The (univariate) ANOVAs, of both representations (see Tables 5.4 and 5.6) reveal that the personality trait neuroticism is of influence on the distinct features, however, only in combination with either the environment, gender, or both. So, personality traits seem to play their role in our emotional experiences and their reflection in our speech and ECG; however, it is a complex interplay of factors, which may be expected to be challenging to unveil.

Various demographic information was gathered on the participants, namely: level of education, age, nationality, and gender. The possible effect of these factors was assessed with preliminary analyses. These analyses have not been reported for reasons of brevity. The preliminary analysis showed the absence of an effect due to the level of education; hence, this factor was excluded from further analysis. However, this lack of effect can be explained by the small variance in level of education between the participants. Age influenced neither the reported emotions nor the physiological signals accompanying them. This is in contrast with some of the literature that states that age is of importance [361]. This is even specifically shown for cardiovascular reactivity on psychological stress [390, 661]. This lack of effect can be explained by the skewed distribution as well as by the limited variance of the age of the participants; see also Section 5.4. The nationality of the participants was heavily skewed towards Dutch: 26 of the 32 participants on which the analyses were conducted did have a Dutch nationality. Therefore, the choice was made to divide the participants having a Dutch and non-Dutch nationality (i.e., consisting of 4 different nationalities). However, this analysis did not reveal any effect on this factor. Nationality was included as a representation of both cultural and ethnical factors. Both these factors have been reported to be of influence on physiological responses in relation to emotions [603]. More than anything else, it should be concluded that this research was not optimized for the assessment of this factor, which explains the absence of any effect. The gender of the 40 participants was perfectly balanced; so, in contrast with level of education and age, for this factor a maximal variance was obtained. In line with the literature, gender was shown to be of effect [392, 661]. However, this effect was marginal and additional research is needed to unveil the exact influence of gender on the relation between biosignals and speech and emotions.

5.7.2 Conclusion

The results of this study show that the three speech measures (i.e., SD F0, $\mathcal{I}$, and $\mathcal{E}$) in combination with only HRV already provide a reliable, robust, and unobtrusive method to reflect user's affective state. Of course, many more features could be derived from both the speech signal [182, 590, 644, 725, 739] and the ECG [14, 85, 105, 481, 538]. However, this was not the aim of this study and also appeared to be unnecessary. The current results are already excellent with 90% explained variance for the VA model (see Section 5.6), which also challenges the claim that personalized processing of such signals is required.

The debate on how to define emotions remains intriguing, as it is so close to our everyday lives. However, for personal and ubiquitous computing technology practical considerations should also be taken into account. The processing scheme introduced in this study enables the unobtrusive assessment of affect. In practice this can be achieved through either sensor networks or wearable sensors; for example, as embedded in electronic textiles [728, 744]. Both of these branches of engineering have flourished since the start of the current century. One can only conclude that this technology is rapidly maturing [10, 138, 438, 508] and, consequently, is applied in a variety of domains; for example, health monitoring [5, 138, 257, 438, 508, 728, 744]. While the underlying technology is becoming both more miniaturized and more robust [257, 438, 508], various probes have been introduced. Generic, ambulatory, wearable ECG systems [414, 513], emotion-aware chairs [18], and digital plasters [728] have been introduced. It seems that sensor networks' and wearable sensors' main drawback is that of many wireless applications, such as your laptop: battery life [5, 257, 438, 728, 744].

All in all, ubiquitous computing, following AI, has to embrace emotion as an essential element in pursuing its next level of development. It is surprising that the combination of speech and biosignals has hardly been used before to unveil people's emotions. Par excellence, this combination of signals has been shown to be suitable for unobtrusive emotion recognition. This having been said, the current study provides a significant leap forward in bringing personal ubiquitous *affective computing* to practice. However, such bold claims should not be made, founded on only one study. Moreover, this study used a subset of the IAPS database and although they again showed to be successful in triggering emotions, this can still be considered as a thin foundation for follow-up research. Therefore, an almost identical study to the one presented in the current chapter will be presented in the next chapter. The study presented in the next chapter differs only with respect to the stimuli used to elicit emotions (cf. [8]). The experiment in the next chapter, Chapter 6, uses the movie fragments introduced in Chapter 3 to elicit emotions.

6

Static versus dynamic stimuli

Abstract

This chapter presents a study that replicated the study presented in the previous chapter, with as its only difference the type of stimuli used to elicit the participants' emotions. So, this study also explores the rare combination of speech, electrocardiogram, and a revised Self-Assessment Mannequin (SAM) to assess people's emotions. Forty people watched movie scenes as were introduced in Chapters 3 and 4 in either an office or a living room environment. Additionally, their scores for the personality traits neuroticism and extroversion and demographic information (i.e., gender, nationality, and level of education) was included in the analysis. The resulting data was analyzed using both basic emotion categories and the valence-arousal model, which enabled a comparison between both representations. It was shown that, when people's gender is taken into account, both heart rate variability (HRV, derived from the ECG) and the standard deviation of the fundamental frequency of speech indicate people's experienced valence and arousal, in parallel. As such, both measures seem to validate each other. However, the explained variance is much lower than on the data of the previous chapter. For the valence-arousal model, the explained variance was reasonable: 43%. In contrast, for the basic emotions, the explained variance was low: 21%. So, in line with the previous chapter, this study also supports in favor of the valence-arousal model and questions the six basic emotions. Further comparison of both studies confirmed that, independent of emotion representation, the bimodal *ASP* approach taken is robust in penetrating emotions. In particular, this is confirmed by the combination of features from the two modalities. The distinct features from speech and HRV reveal a more subtle picture in which the several factors do appear to play their role. An exception is the personality trait extroversion, which seems to be of hardly any influence at all.

This chapter is a thoroughly revised version of:

Broek, E.L. van den, Schut, M.H., Westerink, J.H.D.M., & Tuinenbreijer, K. (2009). Unobtrusive Sensing of Emotions (USE). *Journal of Ambient Intelligence and Smart Environments, 1(3)*, 287–299. *[Thematic Issue: Contribution of Artificial Intelligence to AmI]*

6.1 Introduction

A decade ago, Ducatel, Bogdanowicz, Scapolo, Leijten, and Burgelman [169] expressed a similar concern in "*Scenarios for Ambient Intelligence in* 2010" on behalf of the EU's IST Advisory Group. Two of their key notions were already assessed in the previous chapter and will be further explored in the current chapter: emotion and unobtrusive measurements.The lessons learned in Artificial Intelligence (AI), Cybernetics, psychophysiology, and other disciplines will be taken into account, which will also make it a truly interdisciplinary research.

Before continuing this research, let me take a step back ... in time. Let me cherish remarks made on machine intelligence, either long or not so lang ago. AI pioneer Herbert A. Simon [611] was the first to denote the importance of emotion for AI. Minsky [454, Chapter 16, p. 163] confirmed this by stating: *The question is not whether intelligent machines can have emotions, but whether machines can be intelligent without emotions.* Nevertheless, in practice emotions were mostly ignored in the quest for intelligent machines until Picard [521] introduced the field *affective computing*. Since then, the importance of emotion for AI was slowly acknowledged [455]. However, it needs to be stressed that emotions are not only of crucial importance for true AI but are at least as important for Ambient Intelligence (AmI). This has already been acknowledged by Emile Aarts [1, p. 14]: *Ubiquitous-computing environments should exhibit some form of emotion to make them truly intelligent. To this end, the system's self-adaptive capabilities should detect user moods and react accordingly.*

This chapter continues the quest for ubiquitous *affective computing*. In line with the previous chapter, this study also respects the complexity of emotions as well as the current limitations of unobtrusive physiological measurement. The study reported in this chapter is a replication of the study reported in the previous chapter, except for the stimuli used to elicit the emotions from the participants. To refrain from major redundancies in this monograph, this chapter is a compressed version of the article it originates from.

First, I will introduce the construct emotion (Section 6.2) by taking a step back and briefly discussing the work that served as the foundation of the definition of emotion used in the previous chapter. Next, in Section 6.3, I will denote the aspects on which the study reported in this section deviates from that reported in the previous chapter. Sections 6.4 and 6.5 will describe respectively the preparation of the analysis and its results. Subsequently, in Section 6.6 the current study will be compared with the one presented in the previous chapter. Last, in Section 6.7, I will close this chapter with some final words on the work presented here.

6.2 Emotion

A lengthy debate on the topic of emotion would be justified; however, this falls way beyond the scope of the current chapter, it even falls beyond the scope of this monograph. Hence, in this chapter there will be no overview of the various emotion theories and the levels on which emotions can be described either. Instead, I will take the elaboration on the definition of emotion as provided in the previous chapter as starting point. The definition presented in Chapter 5 was based on recent work by Izard [302] and the comments that followed this work [302]. Carroll E. Izard [302] took the work Kleinginna and Kleinginna [350] as starting point. For a thoroughly composed definition this will be used as a starting point. In line with Izard [302], I will also go back to this work.

Kleinginna and Kleinginna [350] compiled a list of 92 definitions and 9 skeptical statements about emotion. Regrettably, they had to conclude that *psychologists cannot agree on many distinguishing characteristics of emotions*. Therefore, they proposed a working definition: *Emotion is a complex set of interactions among subjective and objective factors, mediated by neural/hormonal systems, which can* (a) *give rise to affective experiences such as feelings of arousal and pleasure / displeasure;* (b) *generate cognitive processes such as emotionally relevant perceptual effects, appraisals, labeling processes;* (c) *activate widespread physiological adjustments to the arousing conditions; and* (d) *lead to behavior that is often, but not always, expressive, goal directed, and adaptive.* I will now adopt this definition as working definition, instead of that of Izard [302], as presented in Chapter 5.

Kleinginna and Kleinginna [350] also addressed the influence of emotions on people's cognitive processes: issues (b) and (d). Hence, emotions by themselves should be taken into account; but, so should their effect on cognitive processes (e.g., attention, visual perception, and memory) and, thereby, our functioning. This emphasizes the importance of taking emotions into account in AmI. Moreover, Kleinginna and Kleinginna [350] addressed the influence of emotions on our physiology. This is nicely in line with the main objective of this monograph: *Affective Signal Processing (ASP): Unraveling the mystery of emotions*.

In line with the frequently adopted circumplex or valence-arousal (VA) model of emotions [372, 443, 535, 647], the definition of Kleinginna and Kleinginna [350] distinguishes arousal and valence (i.e., pleasure / displeasure). The valence-arousal model denotes valence and arousal as two independent bipolar factors that describe emotions. Although the VA model is successful, it has two severe limitations. First, no emotions are identified with high scores, either positive or negative, on either the valence or the arousal axis [372]. Second, the model cannot handle mixed emotions; that is, parallel experience of both positive and negative valence [79] (see also Chapters 3 and 4).

To enable the identification of mixed emotions and provide a suitable processing scheme, the valence-arousal model is sometimes extended [79, 357] (cf. Chapters 3-5). Such an extended VA model incorporates, instead of one bipolar valence dimension, two unipolar valence dimensions: one for positive and one for negative valence. Hence, the extended valence-arousal model incorporates three dimensions, instead of two. This approach was also adopted for the current study.

6.3 Method

This section is compressed similarly to the Method section in Chapter 5. This will facilitate the comparison between both studies, which will have the focus of this chapter. Only those elements that deviate from the work presented in Chapter 5 will be reported here.

The same 40 volunteers participated in this study as did in the study presented in Chapter 5 and they were treated identically. The speech and ECG signals were recorded with the same devices and processed with the same software as has been reported in Chapter 5. The results of the revised, short scale of the Eysenck Personality Questionnaire (EPQ-RSS) [187] completed by the participants in Chapter 5 were also included in the analyses of the current study. For details, see the Methods section of Chapter 5

Instead of looking at IAPS pictures [374, 452], as was the case in the study presented in Chapter 5, dynamic stimuli were chosen [235, 237, 570, 700, 701]. To elicit an emotional response, the participants watched a selection of the six movie scenes that were introduced in Chapters 3 and 4. The movie scenes were presented on the same screen as the IAPS images of Chapter 5 were. Each of the six movie scenes that were shown had a duration of 3 minutes and 18 seconds. After each scene, the participants had 30 seconds to describe the most emotional part of the scene, followed by a rest period of 60 seconds. During these 90 seconds (speaking and resting), a gray screen was shown. The experiment started and finished, displaying a gray screen during 90 seconds.

6.4 Preparation for analysis

The speech signal of two participants was not recorded due to technical problems. The speech signals of two other participants were too noisy. The speech signals of these four participants were excluded from further analyses. With nine participants, either a significant amount of noise was present in their ECG or the signal was completely absent. The ECG signals of these participants were omitted from further processing. In total, 13 corruptions of signals were detected in the signals of 11 participants. So, the recordings of 2 participants suffered from 2 types of noise. Through interpolation, corrections could have been made for the absence of this data. However, this would have decreased the reliability of the analyses done. Therefore, we chose to omit all data for participants for whom prob-

Table 6.1: The six film scenes with the average ratings given by the participants on the positive valence, negative valence, and arousal Likert scales. From the positive and negative valence ratings, three valence categories can be derived: neutral, positive, and negative. Using the scores on arousal, two arousal categories can be determined: low and high

film scene	valence			arousal	
	positive	negative	category	score	category
Color bars	0.13	2.51	neutral	0.49	low
Final Destination	2.59	4.38	neutral	6.54	high
The bear	5.79	0.74	positive	3.49	low
Tarzan	7.31	0.26	positive	4.77	high
Pink flamingos	0.49	7.18	negative	6.00	low‡
Cry freedom	0.56	7.90	negative	7.69	high
Average	2.81	3.83		4.83	

‡ This score is higher than average. Nevertheless, it has been categorized as low. This was done for two reasons: 1) The experienced arousal is low relative to the other film scene with which a negative valence was experienced and 2) This categorization facilitated a balanced design, which enabled the preferred statistical analyses.

lems were encountered with the recordings. This resulted in data from 29 participants that could be analyzed.

Of each of the 29 participants, the sound recorded during the study lasted approximately 25 minutes; however, only the parts in which the participants spoke were of interest. Those parts in which the participants did not speak were automatically omitted from the speech signal processing pipeline.

To assess the emotions experienced by the participants, again the Digital Rating System (DRS) was used, as was introduced in Chapter 5. As in the study presented in Chapter 5, the DRS included three scales: positive valence, negative valence, and arousal, see also Table 6.1. For each film scene, the average ratings were calculated on each of the three scales over all participants. This resulted in a classification of the film scenes into two categories (i.e., high and low) for each of the three scales: positive, negative, and arousal. From these classifications, we derived three categories for valence: positive, negative and neutral. The category neutral denotes neither a positive valence nor a negative valence. In addition, two categories for arousal were derived: high arousal and low arousal. Together, these two categorized dimensions of the VA model depicted six emotion classes. Each of the 6 basic emotions was represented in this research by one film fragment.

Table 6.2: Legend of the factors included in the analyses presented in Section 6.5, particularly in Tables 6.3-6.6.

abbreviation	explanation
V	*v*alence
A	*a*rousal
E	*e*nvironment (or context)
P_N	*p*ersonality trait *n*euroticism
P_E	*p*ersonality trait *e*xtroversion
G	*g*ender

6.5 Results

Here I will discusses the results obtained in this study. First, the considerations taken in the analysis will be presented. Next, in line with the main aims of this study, we will analyze the combined discriminative power of both the combination of features (i.e., repeated measures MANOVA (Wilks' lambda)) and the features separately (i.e., (univariate) repeated measures ANOVA (Huynh-Feldt)). This will be done for two series of analysis, one to assess the dimensional VA model and one to assess six discrete basic emotions. The factors included in the analyses are: environment, the personality traits neuroticism and extroversion, and gender; see also Table 6.2.

6.5.1 Considerations with the analysis

The preliminary analysis of the ECG signal showed that the SD, the variance, and MAD of the heart rate (see Eqs. 5.5 and 5.6) provided similar results. This is in line with both the study presented in Chapter 5 and what is reported in the literature [349, 435]. Also with the current study, we have taken the SD of the R-R intervals as measure for HRV; see also Figure 5.2. Consequently, from this point on, the SD of the R-R intervals will simply be denoted as HRV.

The following factors were also included in the analysis: the environment (i.e., office versus living room), gender, and the two personality traits extroversion and neuroticism. Preliminary analysis showed that the other recorded demographical information (e.g., age; see Section 5.4 of Chapter 5) did not influence the emotional responses of the participants. Hence, this information was excluded from further analyses (cf. Figure 5.2).

As measure of effect size partial eta squared (η^2) will be reported to indicate the proportion of variance accounted for (i.e., a generalization of r/r^2 and R/R^2 in correlation/regression analysis) [737]. The threshold for reporting results is the same as for the study reported in Chapter 5: p $\leq .010$ and, hence, ensures reliable results.

Table 6.3: Results of the repeated measures MANOVA on the valence-arousal (VA) model and its distinct dimensions. The threshold for significance was set to $p \leq .010$.

V	A	E	P_N	P_E	G	Specification of effect
•						F(8,66) = 4.490, $p < .001$, $\eta^2 = .375$
•					•	F(8,66) = 2.850, $p = .009$, $\eta^2 = .257$
	•				•	F(4,15) = 4.999, $p = .009$, $\eta^2 = .571$
•	•					F(8,66) = 6.192, $p < .001$, $\eta^2 = .429$
•	•	•			•	F(8,66) = 3.365, $p = .003$, $\eta^2 = .290$

6.5.2 The (dimensional) valence-arousal (VA) model

In Table 6.3 the results of a repeated measures MANOVA are presented that mapped the four features derived from the speech and ECG model onto the dimensional VA model. Results of both the integral VA model and its two dimensions are presented. Table 6.3 denotes that 43% of the variance of the VA model can be explained with the MANOVA. The results on the distinct dimensions confirm this result with respectively 38% and 57% (when including gender) explained variance for the valence and arousal dimensions.

When the factor gender was included in the analysis, good results were obtained as well. However, including the factors environment and the personality traits neuroticism and extroversion only seemed to add noise to the MANOVA model. The power of the MANOVAs and the explained variance were much lower than when these factors were included. All in all, the assessment of emotions based on the VA model using the four features selected seemed to be robust with respect to influences of environment (or context) and personality traits.

To unravel the influence of each of the four features, a repeated measures ANOVA was executed for each of them. The results of these analyses are presented in Table 6.4. These analyses provide a much more subtle image of the features included in the MANOVA. We will now first discuss the results on the three speech features (i.e., SD F0, intensity $\mathcal{I}$, and energy $\mathcal{E}$) and, subsequently, the ECG feature HRV.

SD F0 was shown to have a good predictive power and was able to explain almost respectively 31% and 42% of all variance for the valence and arousal dimensions of the VA model (when combined with gender information); see Table 6.4. In contrast, for the complete VA model SD F0 had little predictive power. Additional information reduced the power of the model instead of improving it. Effects that exceeded the threshold were not found in any of the analyses for either arousal or valence on the $\mathcal{I}$ or the $\mathcal{E}$.

In the perspective of the VA model, HRV has been reported to indicate both the experienced valence and the experienced arousal. These results were firmly confirmed by the current study; see Table 6.4. Sixty-two percent of all variance in the VA model was explained by solely the HRV of the participants. With 58% explained variance for the valence dimen-

Table 6.4: Results of the repeated measures ANOVAs on the valence-arousal (VA) model and its distinct dimensions. The threshold for significance was set to p $\leq .010$. For the Intensity ($\mathcal{I}$) and Energy ($\mathcal{E}$) of speech no results are reported as none of them exceeded the threshold.

V	A	E	P_N	P_E	G	Specification of effect
SD F0						
•						F(2,36) = 8.186, p = .001, $\eta^2 = .313$
•				•	•	F(2,36) = 7.435, p = .002, $\eta^2 = .292$
	•				•	F(1,18) = 12.863, p < .001, $\eta^2 = .417$
HRV						
•						F(2,36) = 24.937, p < .001, $\eta^2 = .581$
•		•			•	F(2,36) = 10.307, p < .001, $\eta^2 = .364$
	•	•			•	F(1,18) = 16.318, p < .001, $\eta^2 = .475$
	•			•	•	F(1,18) = 8.700, p = .009, $\eta^2 = .326$
•	•					F(2,36) = 29.089, p < .001, $\eta^2 = .618$
•	•	•				F(2,36) = 10.135, p < .001, $\eta^2 = .360$
•	•	•			•	F(2,36) = 15.041, p < .001, $\eta^2 = .455$

sion, this dimension seems to have contributed most to this result. For the arousal dimension effects were only found in interaction with gender and either environment or the personality trait extroversion; see Table 6.4. The explained variance of both the VA model and its dimension valence decreased when additional factors were taken into account. Hence, the complete model as well as one of its dimensions was not sensitive to various interpersonal factors, which is very convenient for many applications.

6.5.3 The six basic emotions

Table 6.5 presents the results of a repeated measures MANOVA that mapped the four features derived from the speech and ECG model onto the six basic emotions. The MANOVA showed an effect of the four features on the six basic emotions, with and without other factors included. Table 6.5 denotes that with the MANOVA 21% of the variance of the six emotions can be explained, which is not much. When other factors are included in the analysis, the power of the analysis declines significantly. Effects of the personality traits neuroticism and extroversion were not found at all. So, the combination of the four features explains

Table 6.5: Results of the repeated measures MANOVA on the six basic emotions. The threshold for significance was set to p $\leq .010$.

E	P_N	P_E	G	Specification of effect
				F(20,360) = 4.764, p < .001, $\eta^2 = .210$
•				F(20,360) = 2.659, p < .001, $\eta^2 = .130$
			•	F(20,360) = 2.704, p < .001, $\eta^2 = .132$
•			•	F(20,360) = 2.362, p < .001, $\eta^2 = .118$

little of the variance between the six basic emotions. However, the model obtained is robust to the various factors included in this study and, hence, as such is robust.

To gain understanding of the influence of the four features, repeated measures ANOVAs were executed for each of them. The results of these analyses are presented in Table 6.6. First, we will discuss the results of the speech features SD F0. Second, the results of the ECG feature HRV will be discussed. In none of the analyses were the speech features intensity $\mathcal{I}$ and energy $\mathcal{E}$ or the personality traits neuroticism and extroversion shown to be of any influence; therefore, these features and factors will not be mentioned further on.

SD F0 by itself was shown to have insufficient power; see Table 6.6. Only when the factors gender with or without the personality trait extroversion were taken into account was an effect found. However, even then the explained variance was limited. The speech features intensity $\mathcal{I}$ and energy $\mathcal{E}$ were shown to have no predictive power at all, neither by themselves nor in combination with other factors; see Table 6.6.

As with the VA model, HRV was also shown to have a high predictive power for the six basic emotions: 57%. When the additional factors were included in the analysis, the explained variance dropped sharply. So, HRV was shown to be robust with respect to environment, personality traits, and gender.

6.5.4 The valence-arousal (VA) model versus basic emotions

When both emotion representations are compared, the following ten main conclusions can be drawn:

- Both emotion representations can handle the variation in participants, even without including additional information such as the environment, personality traits, and gender; see Tables 6.3- 6.6.

Table 6.6: Results of the repeated measures ANOVAs on the six basic emotions. The threshold for significance was set to p $\leq .010$. For the Intensity ($\mathcal{I}$) and Energy ($\mathcal{E}$) of speech no results are reported as none of them exceeded the threshold.

E	P_N	P_E	G	Specification of effect
SD F0				
			•	F(5,90) = 5.501, p $<$ .001, $\eta^2 = .234$
		•	•	F(5,90) = 3.918, p = .003, $\eta^2 = .179$
HRV				
				F(5,90) = 23.772, p $<$ .001, $\eta^2 = .569$
•				F(5,90) = 10.966, p $<$ .001, $\eta^2 = .379$
			•	F(5,90) = 4.128, p = .002, $\eta^2 = .187$
•			•	F(5,90) = 7.456, p $<$ .001, $\eta^2 = .293$

- A high amount of variance can be explained using the VA model: 43%. This is much higher than with the basic emotions: 21% (cf. Tables 6.3 and 6.5).
- Many more effects were found with the VA model compared to the basic emotions as representation for emotions (cf. Tables 6.3 and 6.5 and Tables 6.4 and 6.6).
- The SD F0 was shown to have a good predictive power with both emotion representations; see Tables 6.4 and 6.6.
- The intensity ($\mathcal{I}$) did not contribute to the predictive power of either of the two emotion representations; see Tables 6.4 and 6.6.
- The energy of speech ($\mathcal{E}$) did not contribute to the predictive power of either of the two emotion representations; see Tables 6.4 and 6.6.
- The ECG feature HRV by itself was shown to be able to explain a high amount of variance for both the VA model (62%) and the six basic emotions (57%). For the VA model, this result can be mainly attributed to the valence dimension for which 58% of the variance was explained; see Tables 6.4 and 6.6.
- The personality traits neuroticism and extroversion had no significant influence on the participants' experience of emotions; see Tables 6.3- 6.6.
- Gender was shown to interact with both models, often in combination with other factors. So, gender plays its role; however, what this is cannot be unveiled with the current data.
- Both the VA model and the basic emotions, as treated in this chapter, share many characteristics. This is mainly because a discrete representation of the VA model was used that can distinguish six compounds, similar to the six basic emotions.

This study replicated the study presented in Chapter 5 in its comparison of these two representations. The data of the current study confirmed the findings of the previous chapter that the VA model is most appropriate, as the explained variance is much higher than with the basic emotions: 43% versus 21%. The current results suggest that there is possibly no such thing as a basic emotion (cf. [202]). At best, as with the study in Chapter 5, the discrepancy in explained variance of the present analyses (cf. Tables 6.3-6.6) can be attributed to the variance of the stimuli within one category of one basic emotion. More than anything else this again illustrates the complexity of people and their emotional states. Moreover, this also explains the variation in results reported in the literature, in particular in research that goes from lab to life [674, 680].

6.6 Static versus dynamic stimuli

More than anything else, the current study and its replicate presented in Chapter 5, provide the opportunity to determine the possible influence of the type of stimuli. The current study used dynamic stimuli (i.e., movie scenes) [570], as were already used in the first studies reported in this monograph. In contrast, the study presented in the previous chapter used static IAPS pictures [374, 452]. On the one hand, the movie scenes are claimed to be more ecologically valid [235, 237]. The IAPS images can be questioned with respect to their ecological validity as they present, for example, exaggerated images that are also possibly culturally biased. On the other hand, the IAPS pictures have been validated and have been used frequently and, consequently, have become a sort of standard for emotion elicitation. This having been said, the question remains what (type of) stimuli to use.

When the results of both studies are compared, the first thing that one notices is the difference in the number of effects found. For the VA model and the basic emotions, the results of the study presented in Chapter 5 reported respectively $> 2\times$ and $1.5\times$ the number of results as were reported in the current chapter. For the univariate analysis, a similar trend was shown for the VA model. In contrast, no significant difference in the number of effects was found for the six basic emotions. More interesting than the number of effects is the amount of variance the effects explained. For the VA model, the difference in explained variance between both types of stimuli is enormous: 90% (IAPS pictures) versus 43% (movie scenes). In contrast, for the basic emotions, the difference in explained variance between both types of stimuli was marginal: 18% (IAPS pictures) versus versus 21% (movie scenes). It is noteworthy that these differences are opposite.

The univariate ANOVAs of both studies show a similar trend over both emotion representations. With the VA model many more results were found than with the six emotion categories. This effect seems to have been rather independent of the type of stimuli used. However, more important, the univariate analyses of both studies showed remarkable differences. With the IAPS pictures used as emotion elicitation (see Chapter 5), the speech feature intensity ($\mathcal{I}$) has shown to have a remarkably high discriminative power for the VA model. This result was not confirmed at all in the current study, which employed the movie scenes. Given the fact that everything except the stimuli has been controlled over both studies, this is an astonishing effect. Although it should be noted that neither of the studies revealed any effect of intensity ($\mathcal{I}$) of speech on the six basic emotions.

The study presented here and the one presented in the previous chapter explored more than biomodal emotion elicitation using two distinct types of stimuli. The studies also included various other factors of which it has been posed that they are of influence on people's emotion experience: environment (or context), the personality traits neuroticism and extroversion, and demographic information, most noteworthy gender. In line with what would be expected, these factors were shown to interact significantly. This stresses the need for a holistic approach towards *ASP*, towards a digital human profile, which will be denoted more extensively in the next part of this monograph.

Both studies confirmed that, independent of emotion representation, the bimodal *ASP* approach taken is robust in penetrating emotions. In particular, this is confirmed by the MANOVAs of both studies, see Tables 5.3 and 5.5 as well as Tables 6.3 and 6.5. The ANOVAs reveal a more subtle picture in which the several factors do appear to play their role. An exception is the personality trait extroversion, which seems to be of hardly any influence. Independent of the emotion representation, the personality trait neuroticism had a significant influence on the emotions experienced by the participants when viewing IAPS pictures. Surprisingly, such an effect was not found with emotion elicitation using movie scenes. So, this suggests that the personality trait neuroticism is (also) dependent on the stimuli type and not or not only on the emotions meant to be elicited. Demographic information was shown to be of little value when conducting *ASP*, except for the gender of the participants. Over both studies and both emotion representations, gender was frequently shown to be of influence.

Perhaps the most important conclusion of this one-on-one comparison of the two studies is that, independent of the emotion representation and the type of stimuli used, the speech feature SD F0 and the ECG feature HRV have shown a significant power in penetrating the participants' emotions. Follow-up research should be conducted to see whether or not this effect can be generalized to other types of stimuli and even to other modalities. For the time being, however, the inclusion of both SD F0 and HRV is advised for research and applications that envision emotion-awareness. Additional research should be conducted on the true value of the speech feature intensity ($\mathcal{I}$) and its relation to both emotion representations as used in the two studies discussed here.

6.7 Conclusion

Both the F0 of speech and the HRV can be considered as physiological parameters that can be determined indirectly or at least unobtrusively. This makes them par excellence suitable for AmI purposes. This study and the study reported in the previous chapter were two of the first studies that reported the use of both signals simultaneously to unravel the user's emotional state. To my knowledge, Kim and colleagues [336, 337, 339, 340] are the only ones who have reported on this combination before. The results of this study show that the combination of these measures provides a reliable, robust, and unobtrusive method to penetrate the user's emotional state. Moreover, the signals validate each other. Both HRV and SD F0 seem to indicate influences of experienced valence and arousal in parallel. This also confirmed the findings reported in the previous chapter.

How emotion should be described and modeled remains a topic of debate, despite the work presented in the current and previous chapters. In this chapter, we have adopted the definition of Kleinginna and Kleinginna [350]. However, even in the same decade, various seminal works on emotion were published; for example, Frijda (1986) [208] and Orotony, Clore, and Collins (1988) [502]. Both of these works included their own definition of emotion; for example, Orotony, Clore, and Collins [502, Chapter 1, p. 13 and Chapter 5, p. 191] defined emotions as: *valenced reactions to events, agents, or objects, with their particular nature being determined by the way in which the eliciting situation is construed*. Since the 80s of the previous century, a vast number of books, opinions, and research papers have been published, illustrating the lack of a generally accepted, multidisciplinary theory on emotions. For a concise, more recent overview of the various theories on emotions, we refer to [144, 396, 535, 582].

This chapter closes Part III of this monograph, in which I explored methods and techniques as well as several additional factors to unravel their influence on unveiling emotions. In the next part of the monograph, Part IV, I will present three chapters that explore the feasibility of *affective computing*. In the next chapter, Chapter 7, I will go through the complete signal processing + pattern recognition pipeline, using the data that was also presented in Chapters 3 and 4 and, as such, address the feasibility of emotion-aware systems in a completely different way and will reveal many of its future challenges. Lab research is brought to clinical practice in the two chapters that follow Chapter 7. In Chapter 8 two studies will be presented that explore the feasibility of Computer-Aided Diagnosis (CAD) for mental health care. In these studies, I will employ only the speech signal since direct biosignals were considered to be too obtrusive for the application at hand. After that, in Chapter 9, the complete signal processing + pattern recognition pipeline will be applied on the data derived from the studies discussed in Chapter 8. The resulting analyses can serve as the *ASP* foundation for Computer-Aided Diagnosis (CAD) in mental health care settings.

IV. Towards affective computing

7

Automatic classification of affective signals

Abstract

As we have known for centuries, humans exhibit an electrical profile. This profile is altered by various psychological and physiological processes, which can be measured through biosignals (e.g., electromyography, EMG and electrodermal activity, EDA). These biosignals can reveal our emotions and, as such, can serve as an advanced man-machine interface (MMI) for emotion-aware consumer products. However, such an MMI requires the correct classification of biosignals to emotion classes. This chapter starts with a brief introduction on biosignals for emotion detection. Next, I summarize the research as discussed in Chapters 3 and 4. On this data, several variation of the signal processing + pattern recognition pipeline for *ASP* has been tested, which resulted in a generic framework for automated emotion classification with up to 61.31% correct classification of the 4 emotion classes, without the need of personal profiles. Among various other directives for future research, the results emphasize the need for parallel processing of multiple biosignals.

This chapter is a compressed version of:

Broek, E.L. van den, Lisý, V., Janssen, J.H., Westerink, J.D.H.M., Schut, M.H., and Tuinenbreijer, K. (2010). Affective Man-Machine Interface: Unveiling human emotions through biosignals. In A. Fred, J. Filipe & H. Gamboa (Eds.), *BioMedical Engineering Systems and Technologies (series: Communications in Computer and Information Science, Vol. 52)*, p. 21–47. Berlin/Heidelberg, Germany: Springer-Verlag. *[invited]*

That men are machines (whatever else they may be) has long been suspected; but not till our generation have men fairly felt in concrete just what wonderful psycho-neuro-physical mechanisms they are.

William James (1893; 1842 – 1910)

7.1 Introduction

Despite the early work of William James and others before him, it took more than a century before emotions were widely acknowledged and embraced by science and engineering. However, currently it is generally accepted that emotions cannot be ignored; they influence us, be it consciously or unconsciously, in a wide variety of ways [521]. We are (indeed) *psycho-neuro-physical mechanisms* [312, 440], who both send and perceive biosignals that can be captured; for example, by electromyography (EMG), electrocardiography (ECG), and electrodermal activity (EDA). See Table 1.1 for an overview. These biosignals can reveal a plethora of characteristics of people; for example, workload, attention, and emotions.

Several studies have been conducted in the field of *ASP*, using a broad range of signals, features, and classifiers; see Table 2.4 for an overview. Nonetheless, both the recognition performance and the number of emotions that the classifiers were able to discriminate were disappointing. Moreover, comparing the different studies is problematic because of:

1. The different settings the studies were applied in, ranging from controlled lab studies to real-world testing;
2. The type of emotion triggers used;
3. The number of target states to be discriminated; and
4. The signals and features employed.

All in all, the conclusion has to be that there is a lack of general standards, which results in low prediction accuracy and inconsistent results. However, for *ASP* to come to fruition, it is important to deal with these issues. This illustrates the need for a well documented general framework. In this chapter, I set out to initiate its development, to explore various possibilities, and to apply it on a data set that will be introduced in the next section.

In the pursuit of emotion-aware technology, I will describe our work on the automatic classification of biosignals. Hereby, we will following the complete signal processing + pattern recognition pipeline, as was described in Chapter 1. For an introduction of the statistical techniques that will be employed throughout this chapter, I refer to Appendix A. The data on which the complete signal processing + pattern recognition pipeline will be executed is the data as has been discussed in Chapters 3 and 4. I will refrain from repeating the complete description of this data set here and will only summary it. For the complete description concerning the data set used here, I kindly refer to Chapters 3 and 4.

The remaining chapter is organized as follows: First, in Section 7.3, I will briefly introduce the preprocessing techniques employed. Next, in Section 7.4, the specifications of the pattern recognition techniques will be discussed as well as the classification results they delivered. In Section 7.5, I reflect on my work and critically review it. Finally, in Section 7.6, I will end by drawing the main conclusions.

7.2 Data set

The research in which the data was gathered is already thoroughly documented in Chapters 3 and 4. Therefore, we will only provide a brief summary of it.

The data set concerns the data of 24 subjects who watched movie scenes while affective signals were recorded. In parallel, 4 affective signals were recorded: the EDA and three facial EMG. See Figure 7.2 for samples of the three EMG signals and the EDA signal. These are known to reflect emotions [360]; see also both Table 1.1 and Table 2.4. Regrettably, the affective signal recordings of 3 subjects either failed or were distorted. Hence, the signals of 21 subjects remained for classification purposes.

To elicit emotions with the participants, I selected 8 movie scenes (120 sec. each) for their emotional content. For specifications of these movie scenes, see Chapters 3 and 4. The 8 movie scenes were categorized as being neutral or triggering positive, negative, or mixed (i.e., simultaneous negative and positive; [92]) emotions; hence, 2 movie scenes per emotion category. This categorization was founded on Russell's valence-arousal model [372, 566, 647].

A TMS International Porti $5-16$/ASD system was used for the biosignal recordings and was connected to a PC with TMS Portilab software. Three facial EMGs were recorded: the right corrugator supercilii, the left zygomaticus major, and the left frontalis muscle. The EMG signals were high-pass filtered at 20 Hz, rectified by taking the absolute difference of the two electrodes, and average filtered with a time constant of 0.2 sec. The EDA was recorded using two active skin conductivity electrodes and average filtering with a time constant of about 2 sec. See Figure 7.2 for samples of the three EMG signals and the EDA signal.

7.2.1 Procedure

After the participant was seated, the electrodes were attached and the recording equipment was checked. The 8 movie scenes were presented to the participant in pseudo-random order.

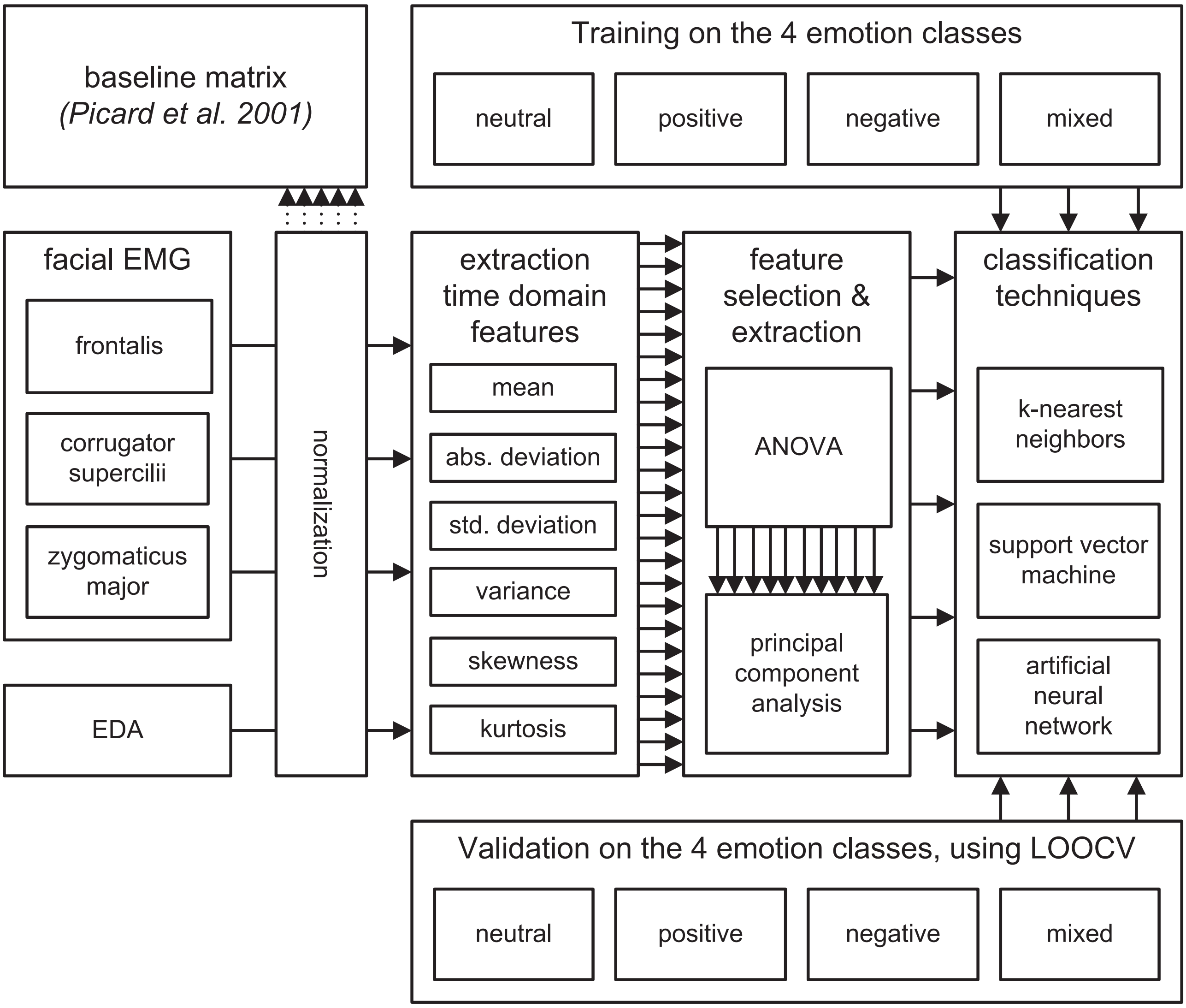


Figure 7.1: The complete processing scheme, as applied in the current research.
Legenda: EMG: electromyography EDA: electrodermal activity; ANOVA of variance; LOOCV: leave-one-out cross validation

A plain blue screen was shown between the scenes for 120 seconds. This assured that the biosignals returned to their baseline level, before the next film fragment was presented.

After the viewing session, the electrodes were removed. Next, the participant answered a few questions regarding the movie scenes watched. To aid their memory, representative print-outs of each fragment were provided.

7.3 Preprocessing

The quest for self-calibrating algorithms for consumer products and for AmI and AI purposes gave some constraints to processing the signals. For example, no advanced filters

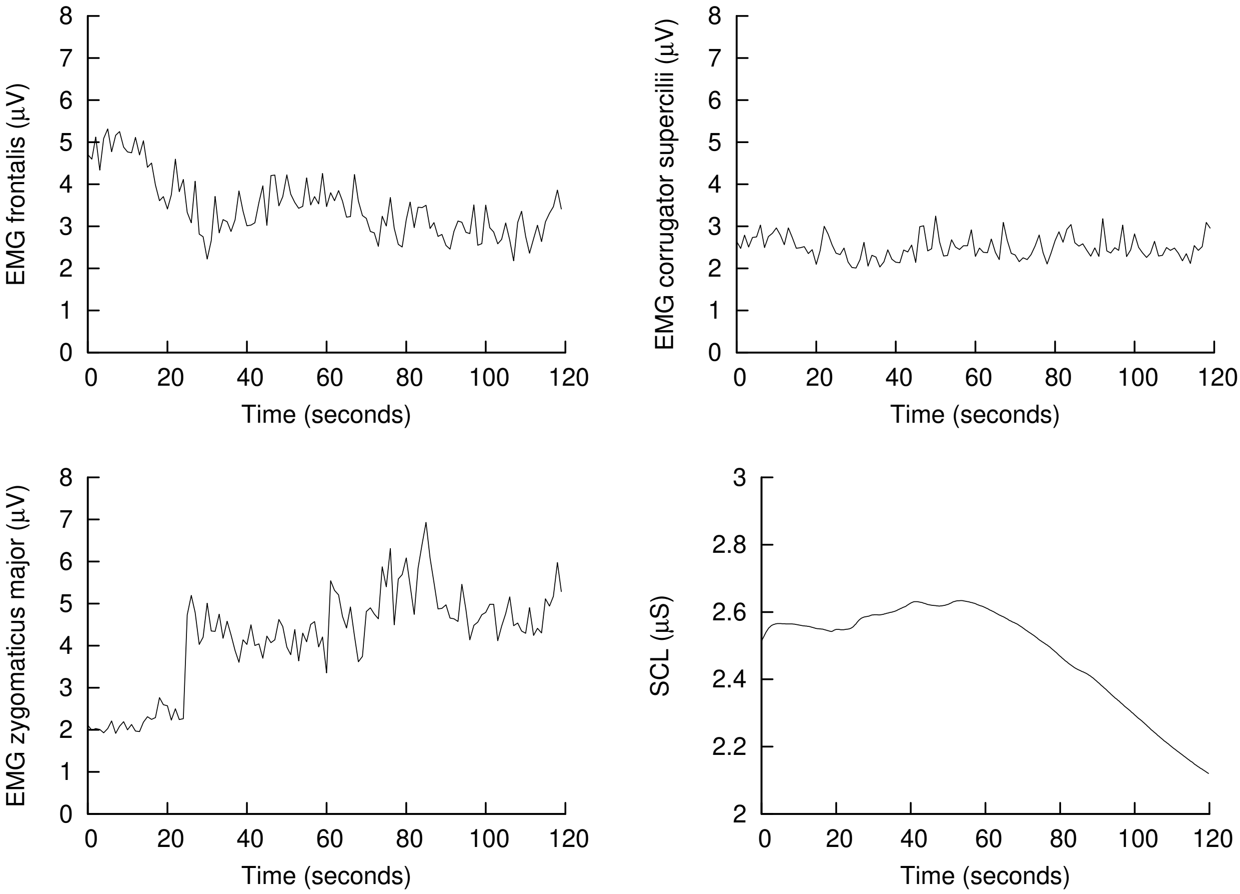


Figure 7.2: Samples of the electromyography (EMG) in μV of the frontalis, the corrugator supercilii, and the zygomaticus major as well as of the electrodermal activity (EDA) in μV, denoted by the skin conductance level (SCL). All these signals were recorded in parallel, with the same person.

should be needed, the algorithms should be noise-resistant, and should (preferably) also be able to handle corrupt data. Therefore, we chose to refrain from advanced preprocessing schemes and applied basic preprocessing. Figure 7.1 presents the complete processing scheme as applied in the current research.

7.3.1 Normalization

Humans are known for their rich variety in all aspects, this is no different for their biosignals. In developing generic classifiers, this required the normalization of the signals. This was expected to boost the performance significantly [541].

For each person, for all his signals, and for all their features separately, the following normalization was applied:

$$x_n = \frac{x_i - \bar{x}}{\sigma},$$

where x_n is the normalized value, x_i the recorded value, $\bar{x}$ the global mean, and σ the standard deviation.

Normalization of data (e.g., signals) was already discussed in Chapters 3 and 4. This has resulted in a variety of normalization functions [48, 62, 457], see also Table 10.4 in the next chapter.

7.3.2 Baseline matrix

In their seminal article, Picard, Vyzas, and Healey (2001) [524] introduced a baseline matrix for processing biosignals for emotion recognition. They suggested that this could tackle problems due to variation both within (e.g., inter day differences) and between participants. Regrettably, Picard et al. (2001) [524] did not provide evidence for its working.

The baseline matrix requires biosignal recordings while people are in a neutral state. Regrettably, such recordings were not available. Alternatively, one of both available neutral movie scenes was chosen; see Chapters 3 and 4.

In line with Picard et al. (2001) [524], the input data was augmented with the baseline values of the same data set. A maximum performance improvement of 1.5% was achieved, using a k-NN classifier. Therefore, the baseline matrix was excluded in the final processing pipeline.

7.3.3 Feature selection

To achieve good classification results with pattern recognition and machine learning, the set of input features is crucial. This is no difference with classifying emotions [191, 443, 680]; see also Section 1.5 and Chapter 2. As was denoted in Chapter 2, biosignals can be processed in the time, frequency, time-frequency, and power domains.

For EMG and EDA signals, the time domain is most often employed for feature extraction; see also Table 1.1. Consequently, we have chosen to explore a range of features from the time domain: mean, absolute deviation, standard deviation (SD), variance, skewness, and kurtosis. Among these are frequently used features (i.e., mean and SD) and rarely used, but promising, features (i.e., skewness and kurtosis) (see Chapters 3 and 4); see also Table 7.1.

To define an optimal set of features, a criterion function should be defined. However, no such criterion function was available in our case. Thus, an exhaustive search in all possible subsets of input features (i.e., 2^{24}) was required to guarantee an optimal set [130]. To limit this enormous search space, an ANOVA-based heuristic search was applied.

For both of the normalizations, we performed feature selection based on ANOVAs. We selected the features with ANOVA $\alpha \leq 0.001$ (see also Appendix A), as this led to the best

Table 7.1: The best feature subsets from the time domain, for k-nearest neighbor (k-NN) classifier with Euclidean metric. They were determined by analysis of variance (ANOVA), using normalization per signal per participant. EDA denotes the electrodermal activity or skin conductance level.

feature	EDA	facial electromyography (EMG) frontalis	corrugator supercilii	zygomaticus
mean				o
absolute deviation				o
standard deviation (SD)		o		o
variance		o		o
skewness	o		o	o
kurtosis			o	

precision. The features selected for each of the biosignals are presented in Table 7.1.

The last step of preprocessing was PCA; see also Appendix A. The improvement after the PCA was small compared to feature selection solely. However, it was positive for normalization; see also Table 7.2. Figure A.1 presents for each set of two emotion classes, of the total of four, a plot denoting the first three principal components. As such, the six resulting plots illustrate the complexity of separating the emotion classes.

7.4 Classification results

This section reports the results of the three classification techniques applied: k-Nearest Neighbors (k-NN), Support Vector Machines (SVM), and Multi-Layer Perceptron (MLP) neural network; see also Appendix A. In all cases, the features extracted from the biosignals were used to classify participants' neutral, positive, negative, or mixed state of emotion; see also Figure A.1. The labels for the emotion classes were provided by the participants, as described in Chapter 3. For the complete processing scheme, we refer to Figure 7.1.

Table 7.2: The recognition precision of the k-nearest neighbors (k-NN) classifier, with $k = 8$ and the Euclidean metric. The influence of three factors is shown: 1) normalization, 2) analysis of variance (ANOVA) feature selection (FS), and 3) Principal Component Analysis (PCA) transform.

normalization	no *fs*	ANOVA *fs* (10 features)	ANOVA *fs* & PCA (5 components)
no	45.54%		
yes	54.07%	60.71%	60.80%

7.4.1 k-Nearest Neighbors (k-NN)

For our experiments, we used MATLAB* and a k-NN implementation, based on SOM Toolbox 2.0†. Besides the classification algorithm described in Appendix A, we used a modified version, more suitable for calculating the recognition rates. Its output was not the resulting class, but a probability of classification to each of the classes. This means that if there is a single winning class, the output is 100% for the winning class and 0% for all the other classes. If there is a tie between multiple classes, the output is divided between them and 0% is provided to the rest. All the recognition rates of the k-NN classifier reported in the current study were obtained by using this modified algorithm.

A correct metric is a crucial part of a k-NN classifier. A variety of metrics provided by the `pdist` function in MATLAB was applied. Different feature subsets appeared to be optimal for different classes. Rani et al. (2006) [541] denoted the same issue in their empirical review (cf. Table 7.1). The results of the best preprocessed input with respect to the four emotion classes (i.e., neutral, positive, negative, and mixed) is 61.31%, with a city block metric and $k = 8$ (cf. Table 7.2).

Probability tables for the different classifications given a known emotion category are quite easy to obtain. They can be derived from confusion matrices of the classifiers by transforming the frequencies to probabilities [355]. Table 7.3 presents the confusion matrix of the k-NN classifier used in this research, with a cityblock metric and $k = 8$.

7.4.2 Support vector machines (SVM)

We have used the MATLAB environment and an SVM and kernel methods (KM) toolbox‡, for experimenting with SVMs. We used input enhanced with the best preprocessing, described in the previous section. It was optimized for the k-NN classifier; however, we expected it to be a good input for more complex classifiers as well including SVM. This assumption was supported by several tests with various normalizations. Hence, the signals were normalized per person, see also Section 7.3. After feature selection, the first 5 principal components from the PCA transformation were selected, see also Appendix A.

The kernel function of SVM characterizes the shapes of possible subsets of inputs classified into one category [586]. Being SVM's similarity measure, the kernel function is the most important part of an SVM; again, see also Appendix A. We applied both a polynomial kernel, with dimensionality d, defined as:

$$k_P(x_i, x^l) = \left(x_i \cdot x^l\right)^d$$

*MATLAB online: http://www.mathworks.com/products/matlab/

†The MATLAB SOM Toolbox 2.0: http://www.cis.hut.fi/projects/somtoolbox

‡The SVM and KM Toolbox: http://asi.insa-rouen.fr/enseignants/~arakotom/toolbox/

Table 7.3: Confusion matrix of the k-NN classifier of EDA and EMG signals for the best reported input preprocessing, with a cityblock metric and $k = 8$.

		real			
		neutral	positive	mixed	negative
classified	neutral	71.43%	19.05%	9.52%	14.29%
	positive	9.52%	57.14%	9.52%	21.43%
	mixed	4.76%	4.76%	64.29%	11.90%
	negative	14.29%	19.05%	16.67%	52.38%

and a Gaussian (or radial basis function) kernel, defined as:

$$k_G(x_i, x^l) = \exp\left(-\frac{|x_i - x^l|^2}{2\sigma^2}\right),$$

where x_i is a feature vector that has to be classified and x^l is a feature vector assigned to a class (i.e., the training sample) [586].

A Gaussian kernel ($\sigma = 0.7$) performed best with 60.71% correct classification. However, a polynomial kernel with $d = 1$ had a similar classification performance (58.93%). All of the results were slightly worse than with the k-NN classifier. Its confusion matrix is similar to that of the k-NN classifier and, hence, is omitted.

7.4.3 Multi-Layer Perceptron (MLP) neural network

We have used a Multi-Layer Perceptron (MLP) trained by a back-propagation algorithm that was implemented in the neural network toolbox of MATLAB; see also Appendix A. It used gradient descent with moment and adaptive training parameters. We have tried to recognize only the inputs that performed best with the k-NN classifier.

In order to assess which topology of ANN was most suitable for the task, we conducted small experiments with both 1 and 2 hidden layers. In both cases, we did try 5 to 16 neurons within each hidden layer. All of the possible $12 + 12 \times 12$ topologies were trained, each with 150 cycles and tested using LOOCV.

The experiments using various network topologies supported the claim that bigger ANN do not always tend to over fit the data [33, 36, 256]. The extra neurons were simply not used in the training process. Consequently, the bigger networks showed good generalization capabilities but did not outperform the smaller ones. An MLP with 1 hidden layer of 12 neurons was shown to be the optimal topology.

An alternative method for stopping the adaptation of the ANN is to use validation data. For this reason, the data set was split into 3 parts: 1 subject for testing, 3 subjects for validation, and 17 subjects for training. The testing subject was completely removed from the training process at the beginning. The network was trained using 17 randomly chosen training subjects. At the end of each training iteration, the network was tested on the 3 validation subjects. This procedure led to a 56.19% correct classification of the four emotion classes. Its confusion matrix is similar to that of the k-NN classifier and, hence, is omitted.

7.4.4 Reflection on the results

Throughout the last decade, various studies have been presented with similar aims. Some of these studies reported good results on the automatic classification of biosignals that should unveil people's emotions; see Table 2.4. For example, Picard et al. (2001) [524] report 81% correct classification on the emotions of one subject [524]. Haag et al. (2004) [246] report 64% − 97% correct classification, using a band function with bandwidth 10% and 20%. This study was conducted on one subject. This study report promising results but also lack the necessary details needed for its replication [246]. More recently, Kim and André (2008) report a recognition accuracy of 95% and 70% for subject-dependent and subject-independent classification. Their study included three subjects [338]. For an exhaustive overview of related studies, see Table 2.4 in Chapter 2.

In comparison with [246, 338, 524], this research incorporated data of a large number of people (i.e., 21), with the aim to develop a generic processing framework. At first glance, with average recognition rates of 60.71% for SVM and 61.31% for k-NN and only 56.19% for ANN, its success is questionable. However, the classification rates differ between the four emotion categories, as is shown in Table 7.3, which presents the confusion matrix of the results of the k-NN classifier. Neutral emotional states are recognized best, with a classification rate of 71.43%. Negative emotional states are the most complex to distinguish from the other three emotion categories, as is marked by the 52.38% correct classification rate. The complexity of separating the four emotion classes from each other is illustrated in Figure A.1.

Taking into consideration the generic processing pipeline (see also Figure 7.1) and the limitations of other comparable research (cf. Table 2.4), the results reported in this chapter should be judged as (at least) reasonably good. Moreover, a broad range of improvements are possible. One of them would be to question the need of identifying specific emotions, using biosignals for MMI. Hence, the use of alternative, rather rough categorizations, as used in the current research, should be further explored.

With pattern recognition and machine learning, preprocessing of the data is crucial. This phase could also be improved for the biosignals used in the current study. First of all, we think that the feature selection based on an ANOVA was not sufficient for more complex classifiers such as neural networks. The ANOVA tests gathered the centers of random distributions that would generate the data of different categories; hereby assuming that their variances were the same. However, a negative result for this test is not enough to decide that a feature did not contain any information. As an alternative for feature selection, the k-NN classifier could be extended by a metric that would weigh the features, instead of omitting the confusing or less informative features.

Taking it all together, the quest towards *affective computing* continues. Although the results presented are good compared to related work, it is hard to estimate whether or not the classification performance is sufficient for embedding of *affective computing* in real-world applications. However, the future is promising with the rapidly increasing number of resources allocated for *affective computing* and the range of improvements that are possible. This assures that the performance on classification of emotions will achieve the necessary further improvements.

7.5 Discussion

This chapter has positioned *men as machines* in the sense that they are *psycho-neuro-physical mechanisms* [312]. It has to be said that this is a far from new position; it has already been known for centuries, although it was rarely exploited in application oriented research. However, in the last decade interest has increased and subareas evolved that utilized this knowledge. This chapter concerns one of them: *affective computing*.

To enable the recognition of these emotions, they had to be classified. Therefore, a brief description of the classification techniques used is provided in Appendix A. Next, a study is introduced in which three EMG signals and people's EDA were measured (see also Figure 7.2), while being exposed to emotion inducing movie scenes; see Section 7.2. See Figure 7.1 for an overview of the processing scheme applied in the current research. Subsequently, preprocessing and the automatic classification of biosignals, using the four emotion categories, were presented in Section 7.3 and Section 7.4.

Also in this research, the differences between participants became apparent. People have different physiological reactions to the same emotions and people experience different emotions with the same stimuli (e.g., music or films). Moreover, these four levels interact [191, 440, 443]. Although our aim was to develop a generic model, one could question whether or not this can be realized. Various attempts have been made to determine people's personal biosignal-profile [338, 440, 524, 541]. However, no generally accepted standard has been developed so far.

In the pursuit of generic *affective computing* processing schemes, the notion of time should be taken into consideration, as I will further elaborate on in Chapter 10. This can help to distinguish between emotions, moods, and personality [31, 565, 676]:

1. Emotion: A short reaction (i.e., a matter of seconds) to the perception of a specific

(external or internal) event, accompanied by mental, behavioral, and physiological changes [191, 680].

2. Mood: A long lasting state, gradually changing, in terms of minutes, hours, or even longer. Moods are experienced without concurrent awareness of their origin and are not object related. Moods do not directly affect actions; however, they do influence our behavior indirectly [191, 222, 680].

3. Personality: People's distinctive traits and behavioral and emotional characteristics. For example, introvert and extrovert people express their emotions in distinct ways. Additionally, self-reports and physiological indicators / biosignals will also be influenced by people's personality trait [123, 670].

With respect to processing the biosignals, the current research could be extended by a more detailed exploration of the time windows; for example, with a span of 10 seconds [191, 443, 679, 680]. Then, data from different time frames could be combined and different, better suitable normalizations could be applied to create new features. For example, information concerning the behavior of the physiological signals could be more informative than only the integral features from a large time window. Studying short time frames could also provide a better understanding of the relation between emotions and their physiological correlates / biosignals, see also Table 1.1.

Other more practical considerations should also be noted. The advances made in wearable computing and sensors facilitates (affective) MMI [220]. In recent years, various prototypes have been developed, which enable the recording of physiological signals [425]. This enables the recordings of various biosignals in parallel. In this way, an even higher probability of correct interpretation can be achieved [191, 443, 676].

Affective MMI can extend consumer products [679]. For example, an music player could sense its listener's emotions and either provide suggestions for other music or automatically adapt its playing list to these emotions. In addition, various other applications have been proposed, mockups have been presented, and implementations have been made. Three examples of these are clothes with wearable computing, games that tweak their behavior and presentation depending on your emotions, and lighting that reacts on or adapts to your mood.

ASP could possibly bring salvation to AI [454, 455, 521, 676]. If we understand and sense emotions, true AI is possibly (and finally) within reach. Current progress in biomedical and electrical engineering provide the means to conduct *affective computing* in an unobtrusive manner and, consequently, gain knowledge about our natural behavior, a prerequisite for modeling it. As AI's natural successor, for AmI [676], even more than for AI, emotion will play a crucial role in making it a success. Since AmI was coined by Emile Aarts [1], this has been widely acknowledged and repeatedly stressed [1, 676]. Also the developments in brain-computer interfacing (BCI) [47, 420, 637] are of interest for *affective computing*. In time, *affective computing* BCI will possibly become within science's reach. Affective BCI, but also BCI in general, could advance AI, AmI, and human-robot interaction. Slowly this is being acknowledged, as is illustrated by a workshop on affective BCI, as was held at the 2009 and 2011 International Conference on Affective Computing and Intelligent Interaction[§]. With affective BCI, again both its scientific foundation and its applications will be of interest.

Without any doubt *affective computing* has a broad range of applications and can help in making various areas more successful. Taking it all together, the results gathered in this research are promising. However, the correct classification rate is below that which is needed for reliable *affective computing* in practice. Providing the range of factors that can be improved, one should expect that the performance can be boosted substantially. That this has not already been achieved is not a good sign; perhaps, some essential mistakes are still being made. One of the mistakes could be the computationally driven approach. A processing scheme that is founded on or at least inspired by knowledge from both biology, in particular physiology, and psychology could possibly be more fruitful . . .

7.6 Conclusions

Affective MMI through biosignals is perhaps the ultimate blend of biomedical engineering, psychophysiology, and AI [22, 714, 716]. However, in its pursuit, various other disciplines (e.g., electrical engineering and psychology) should not be disregarded. In parallel, *affective computing* promotes the quest towards its scientific foundation and screams for its application [191, 443, 680]. As such, it is next generation science and engineering, which could truly bridge the gap between man and machine.

As can be derived from this chapter, various hurdles still have to be taken in the development of a generic, self-calibrating, biosignal-driven classification framework for *affective computing*. The research and the directives denoted here could help in taking some of these hurdles. This can be an important step towards a new, biosignal-driven, era of advanced, *affective computing*. To assess the true value of the signal processing + pattern recognition pipeline presented in this chapter, in the next two chapter research is brought from lab to clinical practice. The feasibility of Computer-Aided Diagnosis (CAD) for mental health care is explored. Chapter 8 presents two studies that employ only the speech signal; direct biosignals were considered to be too obtrusive for the application at hand. In Chapter 9 the complete signal processing + pattern recognition pipeline will be applied on the data derived from the studies discussed in Chapter 8. The resulting analyses can serve as the *ASP* foundation for Computer-Aided Diagnosis (CAD) in mental health care settings.

[§]The IEEE 2009 and 2011 International Conference on Affective Computing and Intelligent Interaction: `http://www.acii2009.nl/` and `http://www.acii2011.org/`

8

Two clinical case studies on bimodal health-related stress assessment

Abstract

This chapter is the first of a set of two chapters that aim towards bringing *affective computing* to practice. As has been denoted in the Introduction, health informatics is on of *ASP*'s application domains. This chapter describes two studies that share the underlying idea that *ASP* can initialize Computer-Aided Diagnosis (CAD) for mental health care. To explore the feasibility of this idea, 25 patients suffering from a Post-Traumatic Stress Disorder (PTSD) participated in both studies. To this date, the treatment of PTSD is a great challenge for therapists. CAD is envisioned to enable objective and unobtrusive stress measurement, provide decision support on whether or not the level of stress is excessive, and, consequently, be able to aid in its treatment. Speech was chosen as an objective, unobtrusive stress indicator, considering that most therapy sessions are already recorded anyway. The two studies concerned a (controlled) stress-provoking storytelling (SPS) and a(n ecologically valid) re-living (RL) study, each consisting of a 'happy' and an 'anxiety triggering' session. The SUD was determined for subjective assessment, which enabled the validation of derived speech features. For both studies, a Linear Regression Model (LRM) was developed, founded on patients' average acoustic profiles. It used five speech features: amplitude, zero crossings, power, high-frequency power, and pitch. From each feature, 13 parameters were derived; hence, in total 65 parameters were calculated. Using the LRM, respectively 83% and 69% of the variance was explained for the SPS and RL study. Moreover, a set of generic speech signal parameters was presented. Together, the models created and parameters identified can serve as the foundation for future CAD tools.

This chapter is an adapted version of:

Broek, E.L. van den, Sluis, F. van der, and Dijkstra, T. (2011). Telling the story and re-living the past: How speech analysis can reveal emotions in post-traumatic stress disorder (PTSD) patients. In J.H.D.M. Westerink, M. Krans, and M. Ouwerkerk (Eds.), *Sensing Emotions: The impact of context on experience measurements (Chapter 10)*, p. 153–180. Series: Philips Research Book Series, Vol. 12. Dordrecht, The Netherlands: Springer Science+Business Media B.V. *[invited]*

No laga duele bieu:	*Let not woes of old*
Skavisábo di nobo.	*enslave you anew.*

– Nydia Ecury –

8.1 Introduction

In our modern society, many people experience stress, sometimes for just a brief moment, at other times for prolonged periods of time. Stress can be defined as a feeling of pressure or tension, caused by influences from the outside world [140, Chapter 6]. It can be accompanied by positive and by negative feelings. It affects our physical state, for instance by increasing our heart rate and blood pressure, and freeing stress hormones like (nor)adrenaline and (nor)epinephrine [359], stimulating autonomic nerve action. Stress may become harmful if it occurs for too long or too frequently, or if it occurs during a traumatic experience. It may, for instance, result in depression, insomnia, or PTSD [178, 365, 547, 562]. To make things even worse, such stress related disorders stigmatize the people suffering from them, which in itself is an additional stressor [563, 564].

Depression cannot always be related to a specific cause, though several contributing factors have been identified: for example, genetic vulnerability and unavoidability of stress [232]. More specifically, certain stressful life events (e.g., job loss, widowhood) can lead to a state of depression. Furthermore, chronic role-related stress is significantly associated with chronically depressed mood [333]. Note that the experience of stress is associated with the onset of depression, and not with the symptoms of depression. Insomnia often has a fairly sudden onset caused by psychological, social, or medical stress [267]. Nevertheless, in some cases, it may develop gradually and without a clear stressor. Insomnia is characterized by sleep deprivation, and associated with increased physiological, cognitive, or emotional arousal in combination with negative conditioning for sleep [9]. Traumas can originate from a range of situations, such as warfare, natural disaster, and interpersonal violence such as sexual, physical, and emotional abuse, intimate partner violence, or collective violence (e.g., a bank robbery) [547]. In such cases, a PTSD may arise, which can be characterized by a series of symptoms and causes [178, 365, 547, 562], summarized in Table 8.1.

8.2 Post-Traumatic Stress Disorder (PTSD)

In our study, we studied the emotions in PTSD patients, who suffered from panic attacks, agoraphobia, and panic disorder with agoraphobia [365, 572].

A panic attack is a discrete period in which there is a sudden onset of intense apprehen-

Table 8.1: Introduction to (the DSM-IV TR [9] criteria for) Post-Traumatic Stress Disorder (PTSD).

Trauma can cause long-term physiological and psychological problems. This has been recognized for centuries. Such suffering (e.g., accompanying a Post-Traumatic Stress Disorder, PTSD), can be characterized in terms of a series of symptoms and causes. Traumas can originate from a range of situations, either short or long lasting; for example, warfare, natural disasters such as earthquakes, interpersonal violence such as sexual, physical, and emotional abuse, intimate partner violence, and collective violence.

Diagnostic criteria as defined by the DSM-IV TR [9] comprise six categories of symptoms, each denoting their various indicators:

1. Exposure of the person to a traumatic event.
2. Persistent reexperience of the traumatic event.
3. Persistent avoidance of stimuli, associated with the trauma, and numbing of general responsiveness (not present before the trauma).
4. Persistent symptoms of increased arousal, not present before the trauma.
5. Duration of the disturbance (symptoms in criteria 2, 3, and 4) is more than one month.
6. The disturbance causes clinically significant distress or impairment in social, occupational, or other important areas of functioning.

Many other symptoms have also been mentioned; for example, weakness, fatigue, loss of willpower, and psychophysiological reactions such as gastrointestinal disturbances. However, these are not included in the DSM-IV TR diagnostic criteria.

Additional diagnostic categories are also suggested for victims of prolonged interpersonal trauma, particularly early in life. These concern problems are related to: 1) regulation of affect and impulses, 2) memory and attention, 3) self-perception, 4) interpersonal relations, 5) somatization, and 6) systems of meaning. Taken together, PTSD includes a broad variety of symptoms and diagnostic criteria. Consequently, the diagnosis is hard to make, as is also the case for various other mental disorders.

sion, fearfulness or terror, often associated with feelings of impending doom. During these Panic Attacks, symptoms such as shortness of breath, palpitations, chest pain or discomfort, choking or smothering sensations, and fear of ‘going crazy’ or losing control are present. The panic attack has a sudden onset and builds rapidly to a peak (usually in 10 minutes or less). Panic attacks can be unexpected (uncued), situationally bound (cued), or situationally predisposed [572]. Agoraphobia is anxiety about, or avoidance of, places or situations from which escape might be difficult (or embarrassing), or in which help may not be available in the event of having a panic attack or panic-like symptoms [572]. Panic disorder with agoraphobia is characterized by both recurrent and unexpected panic attacks, followed by at least one month of persistent concern about having another panic attack, worries about the possible implications or consequences of such attacks, or a significant behavioral change related to these attacks. The frequency and severity of Panic attacks vary widely, but panic disorder as described here has been found in epidemiological studies throughout the world. Panic disorders without and with agoraphobia are diagnosed two to three times as often in women as in men. The age of onset of panic disorders varies considerably, but most typically lies between late adolescence and the mid-thirties. Some individuals may have episodic outbreaks with years of remission in between, and others may have continuous severe symptomatology [572].

Due to its large inter-individual variability and its broad variety of symptoms, the diagnosis of PTSD is hard to make [178, 365, 547, 562]. At the same time, it is clear that an efficient treatment of PTSD requires an objective and early diagnosis of the patients' problems and their therapeutic progress. Assessing the emotional distress of a patient is therefore of the utmost importance. Therapists have developed a range of questionnaires and diagnostic measurement tools for this purpose [353, 572]. Regrettably, these may be experienced as a burden by clients, because it takes the time and willingness of the clients to complete them.

In addition, several other problems arise when a clinician tries to assess the degree of stress in the patient. First, during the appraisal of a stress response, a stressor may not always be seen as stressful enough to be a cause for the mental illness. In other words, although the client may experience it as hugely stressful, the clinician might not always acknowledge it as such. Second, when measuring the response to a stressor, the clinician may rely on introspection and expertise, but these are always to some extent subjective and they also rely on the communicative abilities, truthfulness, and compliance of the client in question. Third, at times it may not be completely clear which (aspect of) the experienced stressor led to the excessive stress response. Finally, the evaluation of the progress in treatment is complicated by its gradualness and relativity.

Given these considerations, it is abundantly clear why researchers have searched for more objective, unobtrusive ways to measure emotions in patient populations. In other words, in addition to standardizing their professional approaches, therapists have sought for new sorts of Computer-Aided Diagnosis (CAD) that are applicable to real-life situations and measure real emotions.

In the following sections, we will first describe both the storytelling and trauma reliving techniques themselves. They provided us with stretches of speech, which we analyzed with respect to a series of signal characteristics to detect emotions. After discussing our speech analysis technique, we will explain how the Subjective Unit of Distress is standardly measured. This will then be followed by a more detailed report of our experimental study. We will end the chapter with an evaluation of our novel approach to stress and emotion measurement.

8.3 Storytelling and reliving the past

As described above, the PTSD patients in our study suffered from Panic Attacks. During and directly after a Panic Attack, there is usually a continuous worrying by the client about a new attack, which induces an acute and almost continuous form of stress. In our main studies, we attempted to mimic such stress in two ways; see also Figure 8.1.

First, in the Stress-Provoking Story (SPS) study, the participants read a stress-provoking or a positive story aloud [685]. Here, storytelling was used as the preferred method to elicit true emotions in the patient. This method allows great methodological control over the invoked emotions, in the sense that every patient reads exactly the same story. The fictive stories were constructed in such a way that they would induce certain relevant emotional associations. Thus, by reading the words and understanding the story line, negative or positive associations could be triggered. The complexity and syntactic structure of the different stories were controlled to exclude the effects of confounding factors. The negative stories were constructed to invoke anxiety, as it is experienced by patients suffering from PTSD. Anxiety is, of course, one of the primary stressful emotions. The positive stories were constructed to invoke a positive feeling of happiness.

Second, in the Re-Living (RL) study, the participants told freely about either their last panic attack or their last joyful occasion [248]. The therapists assured us that real emotions would be triggered in the reliving sessions with PTSD patients, in particular in reliving the last panic attack. As the reliving blocks were expected to have a high impact on the patient's emotional state, a therapist was present for each patient and during all sessions. The two RL sessions were chosen to resemble two phases in therapy: the start and the end of it. Reliving a panic attack resembles the trauma in its full strength, as at the moment of intake of the patient. Telling about the last happy event a patient experienced, resembles a patient who is relaxed or (at least) in a 'normal' emotional condition. This should resemble the end of the therapy sessions, when the PTSD has disappeared or is diminished.

8.4 Emotion detection by means of speech signal analysis

The emotional state that people are in (during telling a story or reliving the past) can be detected by measuring various signals, as has been outlined in Chapter 2. However, more than any other signal, speech suited our purposes as:

1. Speech recordings are completely unobtrusive, see Chapter 2.
2. The communication in therapy sessions is often recorded anyway. Hence, no additional technical effort had to be made on the part of the therapists.
3. Therapy sessions are generally held under controlled conditions in a room shielded

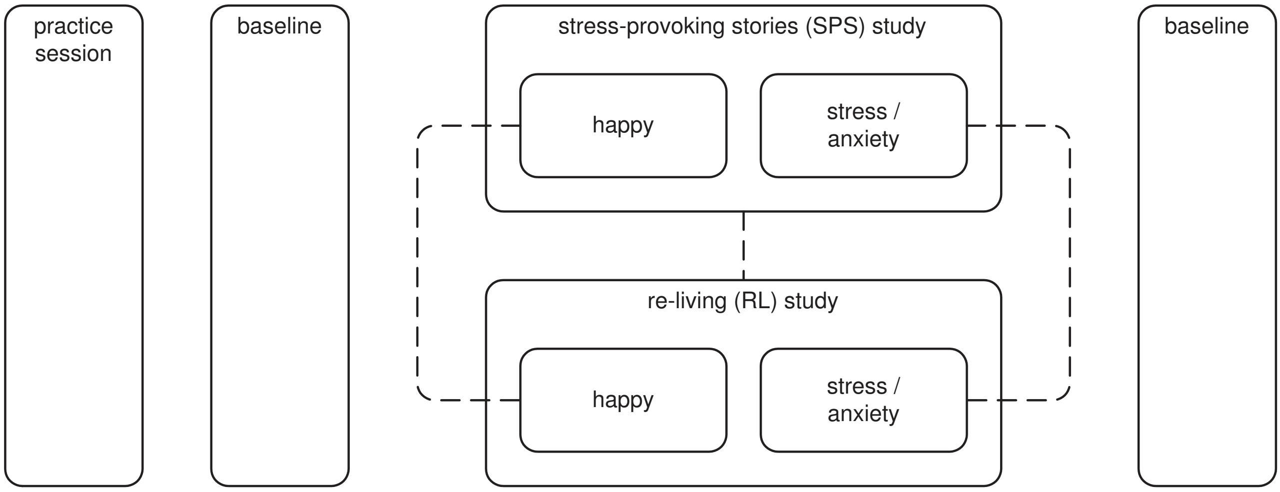


Figure 8.1: Overview of both the design of the research and the relations (dotted lines) investigated. The two studies, SPS and RL, are indicated, each consisting of a happy and a stress/anxiety-inducing session. In addition, baseline measurements were done, before and after the two studies.

from noise. Hence, the degree of speech signal distortion can be expected to be limited.

There is a vast amount of literature on the relationship between speech and emotion, as was already denoted in Chapter 2. Various speech features have been shown to be sensitive to experienced emotions; see Chapter 2 for a concise review. In this research, we extracted five characteristics of speech:

1. the power (or intensity or energy) of the speech signal; for example, see Table 2.2, Chapters 5 and 6, and [131, 469];
2. its fundamental frequency (F0) or pitch, see also Table 2.2, Chapters 5 and 6, and [131, 369, 469, 579, 696];
3. the zero-crossings rate [331, 560];
4. its raw amplitude [469, 579]; and
5. the high-frequency power [27, 131, 469, 560].

All of these characteristics have been considered as useful for the measurement of experience emotions. Moreover, we expect them to be complementary to a high extent.

8.5 The Subjective Unit of Distress (SUD)

To evaluate the quality of our speech analysis, we must compare it to an independent measure of distress. We compared the results of our speech features to those obtained from a standard questionnaire, which measured the Subjective Unit of Distress (SUD). The SUD was introduced by Wolpe in 1958 and has since proven itself to be a reliable measure of a person's emotional state. The SUD is measured by means of a Likert scale that registers the degree of distress a person experiences at a particular moment in time. In our case, we used a linear scale with a range between 0 and 10 on which the experienced degree of distress could be indicated by a dot or cross. The participants in our study were asked to fill in the SUD test once every minute; therefore, it became routine during the experimental session and has not been a stress provoking factor.

8.6 Design and procedure

In our study, 25 female Dutch PTSD patients (mean age: 36; SD: 11.32) participated of their own free will. This group of patients was chosen because we expected that their emotions could be elicited rather easy by way of the two studies, as PTSD patients are sensitive for emotion elicitation methods chosen. Hence, in contrast with traditional emotion elicitation as often conducted in research settings (cf. Chapters 3-6), true emotions were (almost) guaranteed with these studies.

All patients signed an informed consent and all were aware of the tasks included. The experiment began with a practice session, during which the participants learned to speak continuously for longer stretches of time, because during piloting it was noticed that participants had difficulty in doing this. In addition, the practice session offered them the opportunity to become more comfortable with the experimental setting. Next, the main research started, which consisted of two studies and two baseline sessions. The experiment began and ended with the establishment of the baselines, in which speech and SUD were recorded. Between the two baseline blocks, the two studies, the Stress-Provoking Stories (SPS) study and the Re-Living (RL) study, were presented. The two studies were counterbalanced across participants.

The SPS study aimed at triggering two different affective states in the patient. It involved the telling of two stories, which were meant to induce either stress or a neutral feeling. From each of the sessions, three minutes in the middle of the session were used for analysis. The order of the two story sessions was counterbalanced over participants. Both speech and SUD scores (once per minute) were collected. The RL study also involved two sessions of three minutes. In one of these, the patients were asked to re-experience their last panic attack. In the other, the patients were asked to tell about the last happy event they could recall. Again, the order of sessions was counterbalanced over participants. With both studies, problems occurred with one patient. In both cases, the data of this patient were omitted from further analysis. Hence, in both conditions, the data of 24 patients were used for further analysis.

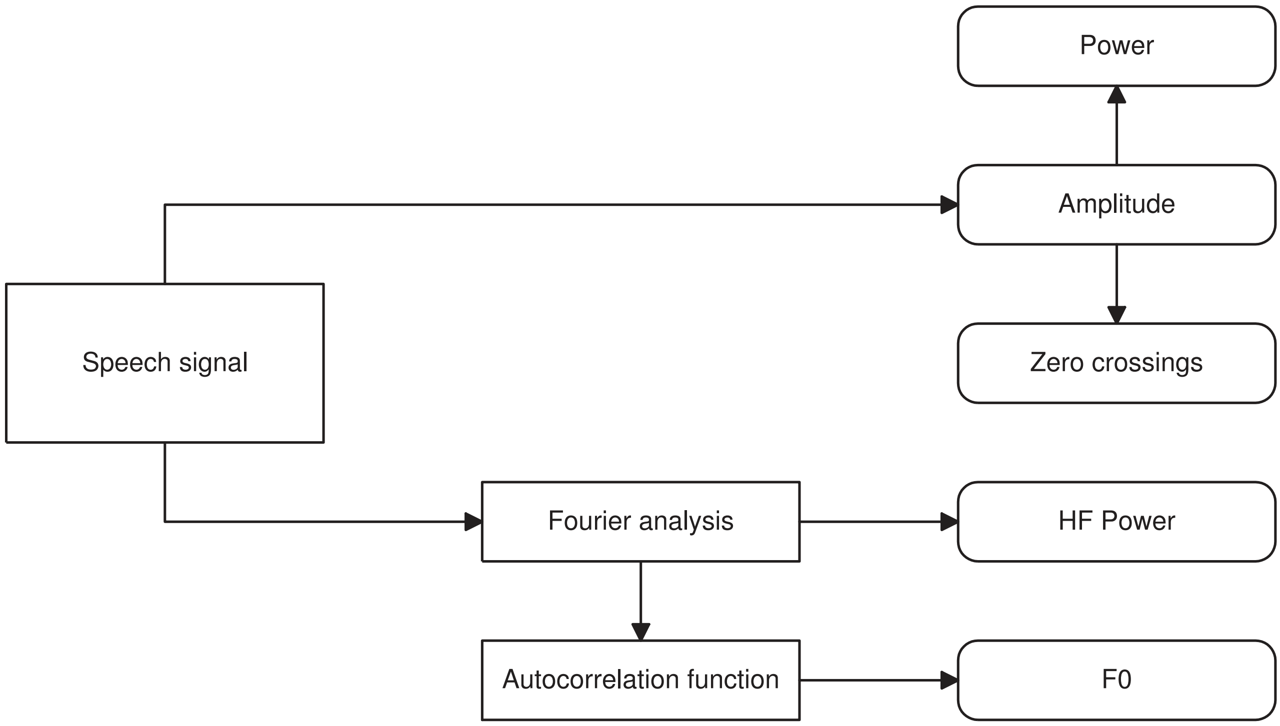


Figure 8.2: Speech signal processing scheme, as applied in this research.
Abbreviations: F0: fundamental frequency, HF: high frequency.

8.7 Features extracted from the speech signal

Recording speech was done using a personal computer, a microphone preamplifier, and a microphone. The sample rate of the recordings was 44.1 kHz, mono channel, with a resolution of 16 bits. All recordings were divided into samples of approximately one minute of speech.

Five features were derived from the samples of recorded speech: raw amplitude, power, zero-crossings, high-frequency power, and fundamental frequency; see also Figure 8.2. Here, we will give a definition of these five features.

The term power is often used interchangeably with energy and intensity. In this chapter, we will follow [432] in using the term power. For a domain $[0, T]$, the power of the speech signal is defined:

$$20\ \log_{10} \frac{1}{\mathrm{P}_0} \sqrt{\frac{1}{T} \int_0^T x^2(t)\ \mathrm{d}t}, \tag{8.1}$$

where the amplitude or sound pressure of the signal is denoted in Pa (Pascal) as $x(t)$ (see also Figure 8.3a) and the auditory threshold P_0 is $2 \cdot 10^{-5}$ Pa [54].

The power of the speech signal is also described as the Sound Pressure Level (SPL),

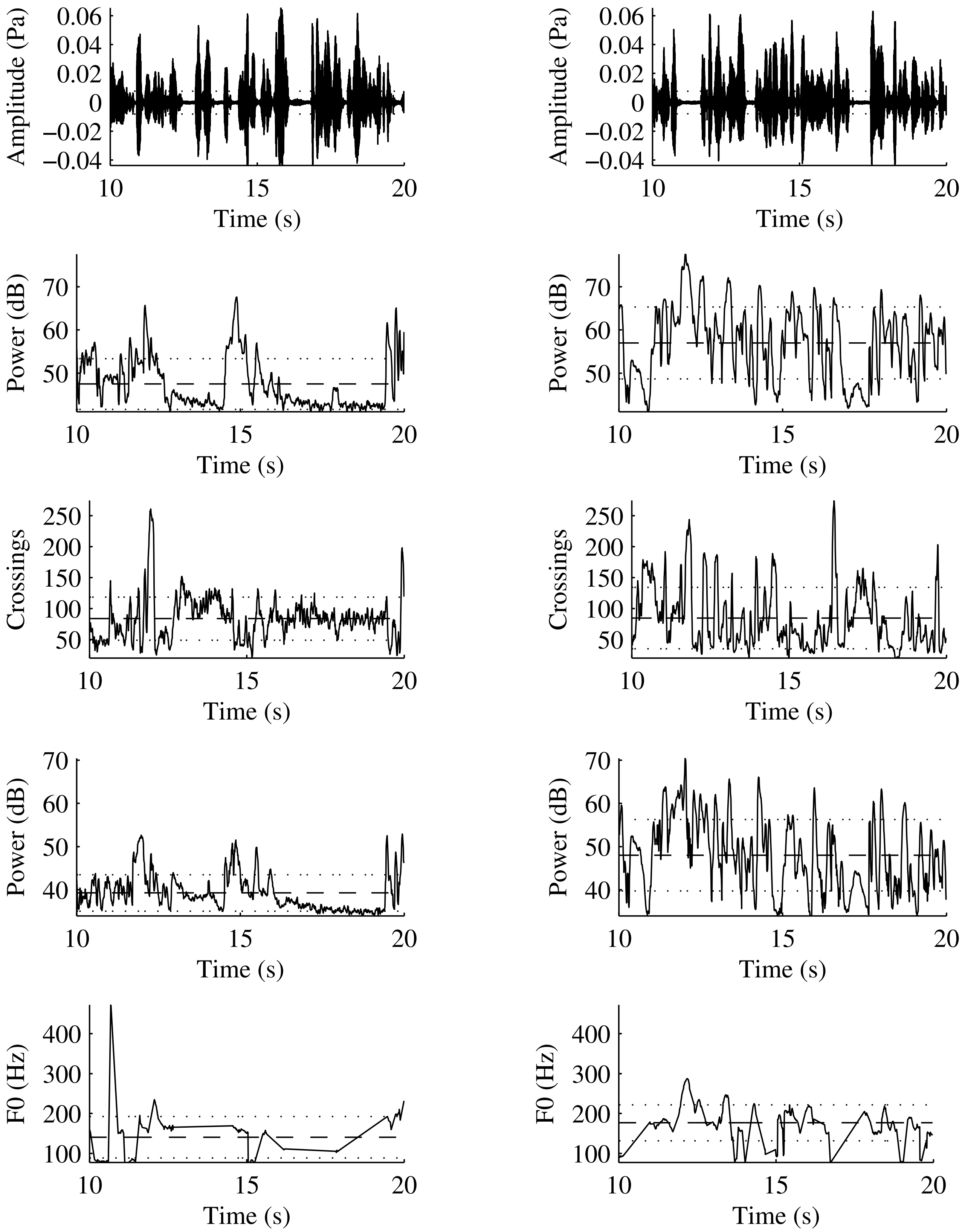


Figure 8.3: A sample of the speech signal features of a PTSD patient from the re-living (RL) study. The dotted lines denote the mean and +/- 1 standard deviation. The patient's SUD scores for this sample were: 9 (left) and 5 (right). Power (dB) (top) denotes the power and the High Frequency (HF) power (dB) (bottom).

calculated by the root mean square of the sound pressure, relative to the auditory threshold P_0 (i.e., in deciBel (dB) (SPL)). Its discrete equivalent is defined as [554]:

$$20\ \log_{10} \frac{1}{\mathrm{P}_0} \sqrt{\frac{1}{N} \sum_{n=0}^{N-1} x^2(n)}, \tag{8.2}$$

where the (sampled) amplitude of the signal is denoted as $x(n)$ in Pa (Pascal) [54]. See Figure 8.3b for an example of the signal power.

The third feature that was computed was the zero-crossings rate of the speech signal. We refrain from defining the continuous model of the zero-crossings rate, since it would require a lengthy introduction and definition (cf. [552]). This falls outside the scope of this chapter.

Zero crossings can be conveniently defined in a discrete manner, through:

$$\frac{1}{N} \sum_{n=1}^{N-1} \mathbb{I}\left\{x(n)x(n-1) < 0\right\}, \tag{8.3}$$

where N is the number of samples of the signal amplitude x. The $\mathbb{I}\left\{\alpha\right\}$ serves as a logical function [331]. An example of this feature is shown in Figure 8.3c. Note that both power and zero-crossings are defined through the signal's amplitude x; see also Figure 8.2, which depicts this relation

The fourth feature that was extracted is the high-frequency power [27]: the power for the domain $[1000, \infty]$, denoted in Hz. In practice, ∞ takes the value $16,000$. To enable this, the signal was first transformed to the frequency domain; see also Figure 8.3d. This is done through a Fourier transform $X(f)$ (see also Figure 8.2), defined as [432]:

$$X(f) = \int_{-\infty}^{\infty} x(t) e^{-j2\pi ft}\ \mathrm{d}t, \tag{8.4}$$

with j representing the $\sqrt{-1}$ operator. Subsequently, the power for the domain $[F_1, F_2]$ is defined as:

$$20\ \log_{10} \sqrt{\frac{1}{F_2 - F_1} \int_{F_1}^{F_2} |X(f)|^2\ \mathrm{d}t}. \tag{8.5}$$

For the implementation of the high-frequency power extraction, the discrete Fourier transform [432] was used:

$$X(m) = \frac{1}{N} \sum_{n=0}^{N-1} x(n) e^{-j2\pi nm/N}, \tag{8.6}$$

with j representing the $\sqrt{-1}$ operator and where m relates to frequency by $f(m) = mf_s/N$. Here, f_s is the sample frequency and N is the number of bins. N typically takes the value of the next power of 2 for the number of samples being analyzed; for example, 640 for a window of 40 msec. sampled at $16,000$ Hz. The power for the domain $[M_1, M_2]$, where $f(M_1) = 1000$ Hz and $f(M_2) = f_s/2$ (i.e., the Nyquist frequency), is defined by:

$$20 \ \log_{10} \frac{1}{\mathrm{P}_0} \sqrt{\frac{1}{M_2 - M_1} \sum_{m=M_1}^{M_2} |X(m)|^2}. \tag{8.7}$$

The fundamental frequency (F0) (or perceived pitch, see Table 2.2) was extracted using an autocorrelation function. The autocorrelation of a signal is the cross-correlation of the signal with itself. The cross-correlation denotes the similarity between two signals, as a function of a time-lag between them. In its continuous form, the autocorrelation r of signal x at time lag τ can be defined as [53]:

$$r_x(\tau) = \int_{-\infty}^{\infty} x(t)x(t+\tau) \ \mathrm{d}t \tag{8.8}$$

In the discrete representation of Eq. 8.8, the autocorrelation R of signal x at time lag m is defined as [607]:

$$R_x(m) = \sum_{n=0}^{N-1} x(n)x(n+m) \tag{8.9}$$

where N is the length of the signal. The autocorrelation is then computed for each time lag m over the domain $M_1 = 0$ and $M_2 = N - 1$. The global maximum of this method is at lag 0. The local maximum beyond 0, lag m_{max}, represents the F0, if its normalized local maximum $R_x(m_{max})/R_x(0)$ (its harmonic strength) is large enough (e.g., $> .45$). The F0 is derived by $1/m_{max}$. See Figure 8.3e for an illustrative output of this method.

Throughout the years, various implementations have been proposed for F0 extraction; for example, [53, 607]. See Table 2.2 for a discussion on speech signal processing and on F0 extraction in particular. In this research, we have adopted the implementation as described in [53]. This implementation applies a fast Fourier transform (see also Eq. 8.4 and Eq. 8.6) to calculate the autocorrelation, as is often done; see [53, 607] and Table2.2. For a more detailed description of this implementation, we refer to [53].

Of all five speech signal features, 13 statistical parameters were derived: $mean$, $median$, standard deviation (std), variance (var), minimum value (min), maximum value (max), range ($max - min$), the quantiles at $10\%(q10)$, $90\%(q90)$, $25\%(q25)$, and $75\%(q75)$, the inter-quantile-range $10\% - 90\%$ ($iqr10, q90 - q10$), and the inter-quantile-range $25\% - 75\%$

($iqr25, q75 - q25$). Except for the feature amplitude, the features and statistical parameters were computed over a time window of 40 msec., using a step length of 10 msec.; that is, computing each feature every 10 msec. over the next 40 msec. of the signal. Hence, in total 65 (i.e., 5×13) parameters were determined from the five speech signal features.

8.8 Results

We analyzed the Stress-Provoking Story study and the Re-Living study separately. The analyses were the same for both studies; with both studies, the SUD scores were reviewed and an acoustic profile was generated.

The acoustic profiles were created with an LRM [262]. An LRM is an optimal linear model of the relationship between one dependent variable (e.g., the SUD) and several independent variables (e.g., the speech features). An LRM typically takes the following form:

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p + \varepsilon,$$

where ε represents unobserved random noise, and p represents the number of predictors (i.e., independent variables x and regression coefficients β). The linear regression equation is the result of a linear regression analysis, which aims to solve the following n equations in an optimal fashion. For more information on LRM, we refer to Appendix A.

It was expected that the acoustic profiles would benefit from a range of parameters derived from the five features, as it is known that various features and their parameters have independent contributions to the speech signal [369]. In order to create a powerful LRM, backward elimination/selection was applied to reduce the number of predictors. With backward elimination/selection, first all relevant features/parameters are added as predictors to the model (the so-called enter method), followed by multiple iterations removing each predictor for which $p < \alpha$ does not hold [155, 262]. In this research, we chose $\alpha = .1$, as the (arbitrary) threshold for determining whether or not a variable had a significant contribution to predicting subjective stress.

The backward elimination/selection stops when for all remaining predictors in the model, $p < \alpha$ is true. As the backward method uses the relative contribution to the *model* as selection criteria, the interdependency of the features is taken into account as well. This makes it a robust method for selecting the most relevant features and their parameters. This is crucial for creating a strong model, because it has been shown that inclusion of too many features can reduce the power of a model [142]. As the general practice of reporting the explained variance of a regression model, R^2, does not take this into account, the adjusted R^2, $\overline{R}^2$ was computed as well. The $\overline{R}^2$ penalizes the addition of extra predictors to the model, and, therefore, is always equal to or lower than R^2.

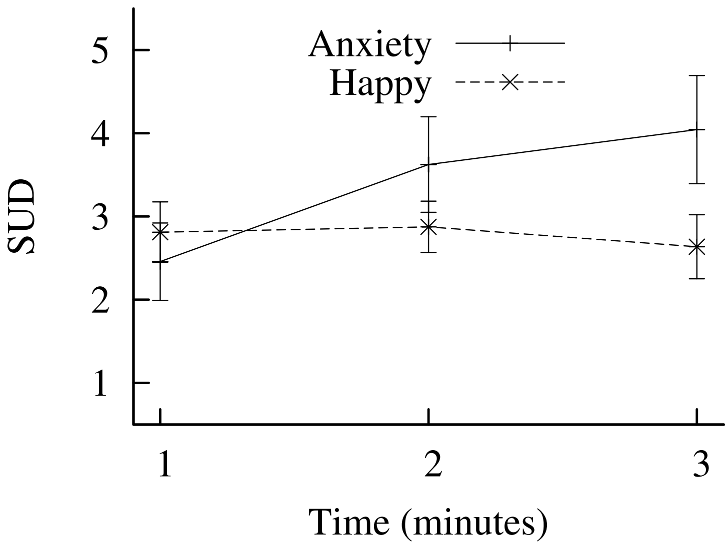


Figure 8.4: Reported stress over time per session (i.e., anxiety triggering and happy) for the Stress-Provoking Stories (SPS) study.

8.8.1 Results of the Stress-Provoking Story (SPS) sessions

Changes with respect to the SUD in the course of the sessions of the SPS study were analyzed first. No main effects of the SPS session (happy or anxious) or measurement moment (first, second, or third minute of storytelling) on the SUD scores were found in an ANOVA, nor did any significant interaction effect between these factors appear. A closer look at the SUD scores in the stress-provoking session showed that the experienced stress reported by the patients increased in the course of storytelling, as indicated by a trend in the ANOVA for the factor measurement moment, $F(2, 67) = 2.59, p < .010$. Figure 8.4 illustrates this trend. In addition, Figure 8.4 shows the confidence intervals, only without variability associated with between-subjects variance (cf. [128]).

Next, a robust acoustic profile was created of the speech characteristics sensitive to stress. This profile was generated after 20 iterations of the backward method, leaving 30 significant predictors explaining 81.00% of variance: $R^2 = .810, \overline{R}^2 = .757, F(30, 109) = 15.447, p < .001$. Before applying the backward method (i.e., before any predictors were removed), 50 predictors explained 82.60% of variance: $R^2 = .826, \overline{R}^2 = .728, F(50, 89) = 8.445, p < .001$. These results indicate that the amount of variance explained through the acoustic profile is high (i.e., $\overline{R}^2 > .75$), as was expected based on the literature [369].

8.8.2 Results of the Re-Living (RL) sessions

Similar to the analyses performed for the SPS sessions, the analyses for the RL sessions start with an ANOVA of the changes in SUD during the course of the sessions. The results

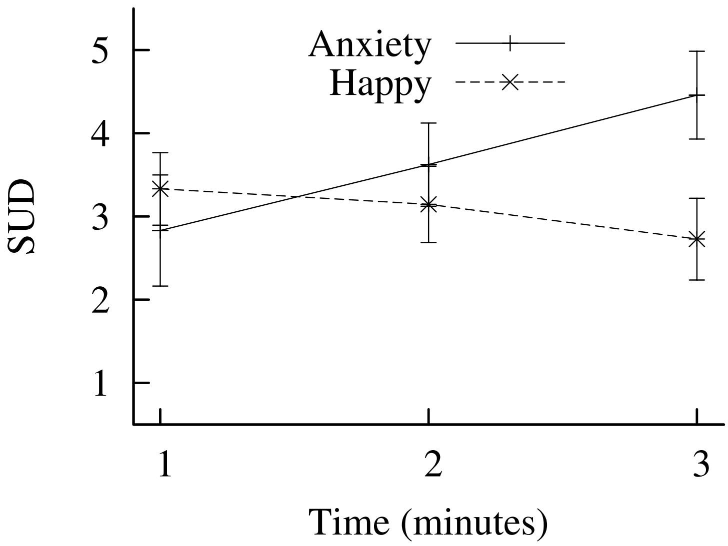


Figure 8.5: Reported stress over time per session (i.e., anxiety triggering and happy) for the Re-Living (RL) study.

were similar to the SPS analyses: no main effects of the RL session (happy or anxious) or time (first, second, or third minute of storytelling) on the SUD scores were found, nor did a significant interaction effect appear. Again, there was a trend in the anxiety triggering condition for patients to report more experienced stress later-on in the course of re-living, as indicated by a trend in the ANOVA for the factor time, $F(2, 69) = 2.69, p < .010$. This trend is also evident in Figure 8.5. Note that Figure 8.5 shows the confidence intervals without between-subjects variance (cf. [128]).

A strong acoustic profile for the RL session was created by means of the speech characteristics that are sensitive to stress. An LRM based upon all relevant features and their parameters (49 predictors) explained 69.10% of variance: $R^2 = .691, \overline{R}^2 = .530, F(49, 94) = 4.29, p < .001$. A smaller LRM, based only on the significant features, used 23 predictors explaining 64.80% of variance: $R^2 = .648, \overline{R}^2 = .584, F(22, 121) = 10.12, p < .001$. These results indicate that, for the RL sessions, the subjectively reported stress could be explained very well, as was expected based on the literature [369]. However, the explained variance was lower than for the SPS sessions.

8.8.2.A Overview of the features

A comparison of the LRM of the RL sessions and the SPS sessions shows that there are 13 shared predictors: *pitch iqr25* and *var*; *amplitude q75*, *var*, and *std*; *power iqr25*, *q25*, and *std*; *zero-crossings q25* and *q10*; *high-frequency power var*, *std*, and *mean*. However, this comparison is misleading due to the role of the interdependency of the predictors in specifying whether or not they have a significant contribution to the estimate. Hence, for

Table 8.2: Correlations between Subjective Unit of Distress (SUD) and the parameters of the five features derived from the speech signal, both for the Re-Living (RL) and the Stress-Provoking Stories (SPS) study.

	Amplitude		Power		ZC		HFP		F0	
Parameter	RL	SPS	RL	SPS	RL	SPS	RL	SPS	RL	SPS
iqr25			-.314¶	-.426¶	-.233§		-.327¶	-.355¶		
q75				-.227§	-.258§			-.196*	-.298¶	-.182*
q25									-.234§	-.244§
iqr10		-.218§	-.358¶	-.428¶	-.197*		-.356¶	-.422¶		
q90		-.209*		-.191*	-.228§			-.189*	-.296¶	-.224§
q10		.225§	.200*	.180*		-.229§	.222§	.168*	-.306¶	-.193*
median					-.180*			-.180*	-.271§	-.202*
min			.223§			-.329¶	.227§			
max					-.192*			-.168*		
range			-.282¶	-.243§	-.179*		-.304¶	-.312¶		
var		-.184*	-.327¶	-.411¶	-.249§		-.317¶	-.384¶		
std		-.202*	-.351¶	-.433¶	-.250§		-.354¶	-.413¶		
mean					-.290¶				-.335¶	-.255§

Levels of significance. $^{*}p < .05$, $^{\S}p < .01$, $^{\P}p < .001$.
Abbreviations. ZC: Zero-Crossings rate, HFP: High-Frequency Power.

a more appropriate comparison, we used a simpler approach; namely, by computing the linear correlation of each feature and its parameters independently of each other for both data sets (i.e., the RL and SPS data). See Table 8.2 for the results.

Table 8.2 shows which predictors are robust for both data sets and which are not; that is, which features show a significant linear correlation for the RL as well as the SPS sessions. The F0 is uniformly robust, namely on its mean and cumulative distribution (q10, q25, median, q75, q90). Power and high-frequency power show similar patterns, though more towards parameters describing the lower part of the cumulative distribution (q10, iqr10, iqr25) and more general statistical parameters used to describe the distribution (std, var, range), only without the mean. There is a perfect similarity between power and high-frequency power in which parameters are relevant for both data sets. The features amplitude and zero-crossings have no parameters relevant for both data sets. Concluding, it seems that F0, power, and high-frequency power, are especially robust features for both data sets.

8.9 Discussion

In this section, we will first briefly discuss the results of both the SPS and RL studies. Next, the results of both studies will be compared to each other. Moreover, the results of both studies will be related to relevant theory.

8.9.1 Stress-Provoking Stories (SPS) study

Stress was successfully induced in and reported by our PTSD patients, using the telling of a carefully created story to induce an affective state (cf. [685]). We were able to define and evaluate an acoustic profile of stress features in speech by comparing speech characteristics to a subjective report of stress. The acoustic profile was shown to explain at best 82.60% of variance of subjectively reported experienced stress.

In interpreting the results, two factors will be differentiated: the experienced and the expressed emotions. In essence, the experienced emotions were targeted by the SUD. Although there was quite some substantial variability in the reported experience, the SUD seemed to have uncovered some expected effects; for example, the stress in the stress inducing story appeared to develop through the course of telling the story. The substantial variability might hint at inter-personal differences which were not evidently expected from the highly standardized stimuli, but which the SUD was able to measure (cf. [367, 368]). Furthermore, another issue can be noted in the experience of the stories; namely, stories develop over time, which implies that a build-up is necessary before an affective state is induced.

As indicated by the explained variance of the acoustic profile, the expressed emotions seem to reflect the experienced emotions very well. In other words, using triangulation through various speech characteristics and the SUD scores indicated that true emotions were indeed triggered and expressed. Hence, although storytelling is only one of many ways to induce emotions, it was particularly useful in creating an emotion-induced speech signal. Contrary to many other methods, of this method the therapists assured us that true emotions would be triggered.

8.9.2 Re-Living (RL) study

Apart from the Stress-Provoking Story (SPS) study, our research included a study in which participants re-lived their traumatic event. As such, this research presents unique data, containing very rare displays of intense, real, emotions; hence, a data set with high ecological validity.

An LRM was generated which explained at best 69.10% of variance in SUD scores, using the RL data set. Although lower than in the SPS study, it is still a very high percentage of explained variance. In interpreting these results, again, we differentiate between the experienced and expressed emotion and used the SUD scores to capture the experienced emotions. The same issues can be denoted as for the SPS study: the SUD scores tended to vary quite substantially across patients, and both showed a build-up in affective state throughout the session. Hence, the experienced emotions varied between patients, which can be expected as the sessions were relatively less standardized [367, 368]; that is, the patients were merely guided in experiencing true emotions. Furthermore, the latter issue is in line with what is known on emotions and their accompanying reactions; that emotions can (indeed) accumulate over time [221, 679].

The expressed emotions are intense displays of emotions; as such, parts of the speech signal even had to be cleaned from non-speech expressions (e.g., crying). Hence, the speech signal clearly reflected emotions. As such, the presented LRM is a rare and clear acoustic profile of true emotions.

8.9.3 Stress-Provoking Stories (SPS) versus Re-Living (RL)

Several differences were found between the studies: the SUD scores for the RL sessions were not significantly higher than for the SPS sessions, and the explained variance of the acoustic profiles was 13.50% lower for the RL study than for the SPS study. Moreover, when comparing the features by their simple linear correlation with the SUD data, it was shown that some features were clearly robust for both studies (i.e., power, high-frequency power, and F0), whereas some were not (i.e., amplitude and zero-crossings rate). In sum, there were 22 parameters (of which 17 were in the amplitude and zero-crossings rate features) which worked for only one of the data sets and 18 parameters which worked for both data sets. The robust parameters could be grouped into specific meaningful parts of the features: for the F0 its mean and cumulative distribution (q10, q25, median, q75, q90), and for power and high-frequency power their lower part of the cumulative distribution (q10, iqr10, iqr25) and more general statistical parameters used to describe the variation of the distribution (std, var, range). Concluding, there were substantial similarities as well as differences between the studies, which will be discussed next.

Considering the experienced emotions, the results were counter-intuitive: the reported stress was not significantly higher in the RL study than in the SPS study. Hence, either the experience was indeed not different from the SPS studies, or introspection is fallible. There were, of course, differences in the experienced emotions between the studies (i.e, the stimuli were different; cf. [8]). Storytelling was used as a highly standardized laboratory method, whereas the re-living sessions were indeed closer to the patients' experiences. Moreover, this view is also supported by the differences between the acoustic profiles and, by qualitative judgements of the patients' psychiatrists also present during the studies. Hence, this would indicate that the SUD scores were a non-perfect mapping on the truly experienced stress. Even if the actual experienced emotions differed between studies, this should not have caused any differences, as the SUD was designed to query this exact experience. Hence, introspection seems to be fallible. Of course, the problems with introspection are not new; tackling them is one of the core motivations for this study. Moreover, we analyzed the SUD scores as an interval scale, an assumption that might not be correct.

The differences between the SPS and the RL study can also be explained by the notion of emotion specifity or cognitive versus emotional stress [367, 415, 563, 564]. Cognitive stress is defined as the information processing load placed on the human operator while performing a task. Emotional stress is the psychological and physiological arousal due to emotions triggered before or during a task. Both the research setting and the therapeutic setting could have caused cognitive stress; so, this would not discriminate between the two studies. However, the cognitive stress possibly had a higher impact on the speech signal obtained with the SPS study than on that obtained with the RL study, where emotional stress was dominant.

Part of the explanation may also lie with the expression of emotions. Already more than a century ago [439], the differentiation between emotional and emotive communication was noted. Emotional communication is a type of spontaneous, unintentional leakage or bursting out of emotion in speech. In contrast, emotive communication has no automatic or necessary relation to "real" inner affective states. Emotive communication can be defined as strategic signaling of affective information in speaking to interaction partners. It uses signal patterns that differ strongly from spontaneous, emotional expressions and can be both intentionally and unintentionally accessed [27]. It is plausible that in the RL study relatively more emotional communication took place, while emotional expressions in the SPS study were based more on features of emotive communication.

When the differences in results between the SPS and the RL study are explained in terms of the distinction between emotional and emotive communication [27, 334, 439], interesting conclusions can be drawn. The intersection of the parameter sets of both studies should then reflect the aspects of the speech signal that are used with emotional communication. The RL study triggered "real" emotions and in the SPS study probably "real" emotions were also revealed in addition to the emotive communication. Consequently, the parameters unique for the SPS study should reflect characteristics of the speech signal that represents emotive communication. Additionally, the parameters unique for the RL study should reflect characteristics of the speech signal that represents emotional communication. Further research investigating this hypothesis is desirable.

Having discussed hypotheses based on both the distinction between cognitive and emotional stress and the theory on emotive and emotional communication, both notions should also be taken together. Communication as expressed with emotional stress [367, 415, 563, 564] and emotional communication [27, 439] could point to the same underlying construct of emotionally loaded communication. However, this does not hold for cognitive stress [367, 415, 563, 564] and emotive communication [27, 334, 439]. It is possible that both cognitive stress and emotive communication played a significant role in the SPS study. This would then involve a complex, unknown interaction. An initial description could include the intersection of both parameter sets that could reveal the aspects of the speech signal that are used both with emotional and emotive communication. This could reflect the cognitive stress experienced. Consequently, the parameters unique for the SPS study would reflect the interaction between emotive communication and cognitive stress. The parameters unique for the RL study would then reflect "real" stress, as was meant to be. If true, this hypothesis would have substantial impact on emotion literature. Therefore, a substantial amount of follow-up research should conducted with the aim to unravel the relation between these theoretical constructs as well as their relation to speech.

8.10 Reflection: Methodological issues and suggestions

The design of this research makes it unique in its kind; see also Figure 8.1. Two studies were conducted, which were alike and at the same time completely different. The Stress-Provoking Stories (SPS) study comprised a controlled experimental method intended to elicit both stress and more happy feelings. Within the Re-Living (RL) study, true emotions linked to personally experienced situations were facilitated. The same patients participated in both studies. The studies were executed sequentially, in a counterbalanced order.

A question which is often posed is whether 'true' emotions can be triggered in controlled research environments. Moreover, if emotions can be triggered in controlled research, how do they relate to emotions experienced in everyday life? Is it only the intensity in which they differ or do different processes underly real-life situations? These questions are hard to answer solely based on a review of the literature. Problems arise when one compares empirical studies.

The validity of the current research is high. Content validity is high as a) the research aimed at a specific group of patients, b) the SUD as well as the speech signal features and their parameters were chosen with care, all denoted repeatedly in the literature; see also Section 8.4, and c) the SUD in combination with the speech signal features chosen provide a complete image of the patients' emotional states, as has been shown. Criteria-related validity is also high as the speech signal was the preferred measurement, being robust and unobtrusive. Moreover, we were able to record emotions real-time. The SUD was provided each minute, which can also be considered as accurate, given the context. The construct validity is limited since for both stress and emotions various definitions exist and no general consensus is present. Moreover, no relations have been drawn between emotion, stress, psychological changes, physiological changes, and the speech signal. The ecological validity is high, at least for one of the studies. For the other study the ecological validity is limited, as illustrated by the difference in results between both studies.

The principle of triangulation is applied [273]; that is, multiple operationalizations of constructs were used. The distinct speech signal features could be validated against each other and against the SUD. Extrapolations were made using the data sets of both studies and a set of common discriminating speech features were identified. Moreover, the SUD was used as ground truth. However, this required introspection of the patients, which is generally not considered as the most reliable measure.

This research has used one signal; hence, no multi-model integration of signals has been applied. However, for both studies, the features and their parameters were all integrated in one LRM. Additional other signals were omitted on purpose since they could contaminate the ecological validity of the research, as they would interfere with the actual tasks the patients had to perform.

CAD should be able to function in a setting such as in which this research was conducted; hence, having the same characteristics. In general, these are average office settings. Within reason, the speech signal processing scheme 8.2 should be able to handle changing characteristics of an office, which could influence the room's acoustics. However, there are no indications for any problems that could occur as a results of this.

8.11 Conclusions

This chapter has presented two studies in which the same PTSD patients participated. This provided us with two unique data sets. This has revealed interesting common denominators as well as differences between both studies, which are of concern for several theoretical frameworks, as was denoted in the previous section. Moreover, a thorough discussion has been presented, in two phases. First, the results of both studies were discussed and, subsequently, related to each other. Second, a range of aspects concerning the complete research were discussed. This emphasized the strength of the research presented and also provided interesting pointers for follow-up research.

A Linear Regression Model (LRM) was developed, derived from the data of each of the studies. These LRMs explained respectively 83% of the variance for the SPS study and 69% of the variance for the RL study, which are both high. Founded on the results of both studies, a set of generic features has been defined; see also Table 8.2. This set could serve as the foundation for the development of models that enable stress identification in a robust and generic manner.

It would also be of interest to apply such a model on patients suffering from other related psychiatric disorders, such as depression [9, 333] and insomnia [9, 267]. Probably, even for less related psychiatric disorders, the current approach would be a good starting point. In such a case, the general framework and speech signal processing scheme (see Figure 8.2), as presented in this chapter, could be employed. Most likely, only the set of parameters used for the LRM would have to be tailored to the specific disorders.

The speech signal processing approach used in this research could also be linked to approaches that measure physiological responsiveness of PTSD in other ways; for example, using biosignals or computer vision techniques (see Chapter 2). This would facilitate a triangulation of the construct under investigation, providing even more reliable results [680]. Furthermore, more specific analyses can be conducted; for example, in terms of either the valence and arousal model or discrete emotion categories [680] (cf. Chapters 6 and 5). However, it also has its disadvantages, as discussed in the previous section.

Taken together, an important and significant step has been made towards CAD for treatment of patients suffering from a PTSD in particular and stress-related psychiatric disorders in general. Through the design of the research, it was made sure that "real" emotions were measured. Subsequently, their objective measurement through speech signal processing was shown to be feasible. Models were constructed, founded on a selection from 65 parameters of five speech features. With up to 83% explained variance, the models were shown to provide reliable, robust classification of stress. As such, the foundation was developed for an objective, easily usable, unobtrusive, and powerful CAD.

With this chapter, theory has been brought to (clinical) practice. Through two studies, it is shown how rich speech is as an indirect biosignal. As such, it can be valuable even without other biosignals added to it. This provides us with an indirect completely unobtrusive biosignal on which models were build that can serve as an expert system in psychiatric practice. The next chapter has little in common with the current chapter except that it also explores the feasibility of building emotion-aware systems. In the current chapter, the research presented in this chapter will be fed to the signal processing + pattern recognition pipeline, aw was introduced in Section 1.5 (see also Figure 1.2) and already employed in Chapter 7. Again, a range of signal processing and machine learning techniques will be presented, which will bring us close to the envisioned emotion-aware systems: *ASP*-based Computer-Aided Diagnosis (CAD) for mental health care.

9

Cross-validation of bimodal health-related stress assessment

Abstract

This chapter is the second of the set of two chapters that aim towards bringing *affective computing* to practice. The previous chapter described two studies with the aim to employ *ASP* to initialize Computer-Aided Diagnosis (CAD) for mental health care. This chapter continues this endeavor with the data of these two studies. Two instruments were chosen to assess the stress level of the patients at various point in time during therapy: i) speech, used as an objective and ubiquitous stress indicator, and ii) the Subjective Unit of Distress (SUD), a clinically validated Likert scale. In total, 13 statistical parameters were derived from each of 5 speech features: amplitude, zero crossings, power, high-frequency power, and pitch. To model the emotional state of the patients, 28 parameters were selected from this set by means of a linear regression model presented in the previous chapter. Subsequently, this representation was compressed into 11 principal components. The SUD and speech model were cross-validated, using 3 machine learning techniques (i.e., k Nearest Neighbors, Support Vector Machines, and Multi-Layer Perceptron neural network). Between 90% (2 SUD levels) and 39% (10 SUD levels) correct classification was achieved. The two sessions could be discriminated in 89% (for ST) and 77% (for RL) of the cases. This report fills a gap between laboratory and clinical studies, as presented in the previous chapter, and its results emphasize the usefulness of Computer Aided Diagnostics (CAD) for mental health care.

This chapter is based on:

Broek, E.L. van den, Sluis, F. van der, and Dijkstra, T. Cross-validation of bimodal health-related stress assessment. *Personal and Ubiquitous Computing*. *[in press]*

9.1 Introduction

Both researchers and clinicians have searched for a long time for more objective, ubiquitous ways to measure stress-like phenomena in (patient) populations [17, 402, 483], involving, for instance, the use of virtual reality technology and biofeedback [506]. In parallel, ubiquitous computing has gradually emerged as an increasingly important paradigm over the last two decades. Excellent up-to-date state-of-the-art overviews on ubiquitous computing are provided by Krumm [363] and Friedewald and Raabe [207]. In addition to the notion of computing itself, intelligence and emotion quickly became important terms in ubiquitous computing. However, as shown repeatedly over 15 years, modeling these is still a bridge too far for current state-of-the-art science and technology (cf. [521]). Even last year, it was remarked that *"pervasive healthcare research in the field of stress prevention is still at an exploratory stage"* [17, p. 70]. Despite such scepticism, the ability to reliably and unobtrusively recognize stress in people might make a more realistic (and consequently better) starting point than either *affective computing* or modeling general (human) intelligence.

In this research, the same 13 statistical parameters were derived from the five speech signal features as were in the previous chapter, namely: $mean$, $median$, standard deviation (std), variance (var), minimum value (min), maximum value (max), range ($max - min$), the quantiles at $10\%(q10)$, $90\%(q90)$, $25\%(q25)$, and $75\%(q75)$, the inter-quantile-range $10\% - 90\%$ ($iqr10, q90 - q10$), and the inter-quantile-range $25\% - 75\%$ ($iqr25, q75 - q25$). The features and statistical parameters were computed over a time window of 40 ms, using a step length of 10 ms, as was done in Chapter 8. However, in this research two variations of amplitude will be reported, instead of one as was done in Chapter 8. The amplitude was determined as both the mean amplitude per window of 40 ms (reported as amplitude(window)) and as calculated over the full signal (reported as amplitude(full)). So, in total, not 65 but $6 \times 13 = 78$ parameters were determined on the basis of the speech signal features.

9.2 Speech signal processing

The quest for self-calibrating algorithms for consumer products, either personalized or ubiquitous, provided some constraints on speech signal processing. For example, no advanced filters should be needed, the algorithms should be noise-resistant and they should (preferably) be able to handle corrupt data.

We therefore only applied some basic preprocessing to the speech signal: outlier removal, data normalization, and parameter derivation from the complete set of features. The first and last aspects require some clarification.

9.2.1 Outlier removal

The same procedure for outlier removal was executed on all speech features. It was based on the inter-quartile range IQR, defined as:

$$IQR = Q_3 - Q_1, \tag{9.1}$$

with Q_1 being the 25th percentile and Q_3 being the 75th percentile. Subsequently, x was considered to be a normal data point if and only if:

$$Q_1 - 3IQR < x < Q_3 + 3IQR. \tag{9.2}$$

All data points that did not satisfy Eq. 9.2 were removed from the data set.

9.2.2 Parameter selection

To achieve good classification results with pattern recognition and machine learning methods, the set of selected input features is crucial. The same holds for classifying stress and emotions. However, there is no criterion function available for our data to define an optimal set of features. As a consequence, an exhaustive search in all possible subsets of input parameters (i.e., 2^{78}) was required to guarantee an optimal set [130]. To limit this enormous search space, a LRM-based heuristic search was applied, using $\alpha \leq 0.1$, which can be considered as a soft threshold.

Similar as in the previous chapter, an LRM was generated using all available data. The process started with the full set of parameters and, then, reducing it in 32 iterations by means of backward removal to a set of 28 parameters. The final model is shown in Table 9.2.2. The parameters in Table 9.2.2 are considered to be the optimal set of parameters and used further on in the processing pipeline.

The LRM in Table 9.2.2 explained 59.2% ($R^2 = .592, F(28, 351) = 18.223, p < .001$) of the variance. This amount of explained variance is low in comparison to previously reported results [714, Chapter 10]: an LRM model based only on the story telling (ST) conditions explained 81.00% of variance: $R^2 = .810, \overline{R}^2 = .757$, $F(30, 109) = 15.447, p < .001$, whereas a model based only on the re-living (RL) conditions explained 64.80% of variance: $R^2 = .648, \overline{R}^2 = .584$, $F(22, 121) = 10.12, p < .001$. The difference in explained variance can be attributed to the selection of data on which the LRMs were based. First, the speech data in the ST conditions are cleaner than in the RL conditions, yielding better models for the ST data. Second, the baseline conditions have normal levels of variance in the speech parameters, but almost no variance in SUD responses; almost no stress was reported in the baseline conditions. This combination of points led to more noise

Table 9.1: Standardized regression coefficients β of a LRM predicting the SUD using speech parameters. HF denotes High-Frequency.

Parameters	Features					
	Amplitude (full)	Amplitude (window)	Power	Zero-Crossings	HF Energy	Pitch
mean				-1.90***	-2.04*	-0.75**
median			1.57***			
std			2.32**		-1.52*	
var	0.83***	-0.22*	-1.71	0.67***	2.04***	0.10
min					0.61**	
max	-0.12*					
range						
q10				-0.26**	0.70*	
q25	1.23***		-2.14***	0.97***	1.39**	0.66**
q75	1.54***			0.63***		
q90	-1.68***	0.78***		0.53***		
iqr10						
iqr25			-1.16***			0.20*

Levels of significance: *** $p \leq .001$, ** $p \leq .01$, *$p \leq .05$. For all other parameters: $p \leq .10$.

in the relation between SUD and speech parameters. However, because the LRM in Table 9.2.2 is used for preprocessing and not as an end result, the LRM had to be applicable to the full data set; hence, it was based on all available data.

9.2.3 Dimensionality Reduction

A Principal Component Analysis (PCA) can be used to further reduce the dimensionality of the set of speech signal parameters, while preserving its variation as much as possible; see also Appendix A. The speech parameters are transformed to a new set of uncorrelated but ordered variables: the principal components $\alpha \cdot x$. The first principal component represents, as well as possible, the variance of the original parameters. Each succeeding component represents the remaining variance, as well as possible. Once the vectors α are obtained, a transformation can map all data x onto its principal n components:

$$x \rightarrow (\alpha_0 \cdot x, \alpha_1 \cdot x, \ldots, \alpha_{n-1} \cdot x)\,.$$

Out of the 78 parameters selected by means of the LRM on the basis of the 5 speech signal features, we selected 28. These 28 parameters were fed to the PCA transformation. Subsequently, the first 11 principal components from the PCA transformation were selected, covering 95% of the variance in the data. These principal components served as input for the classifiers that will be introduced next.

9.3 Classification techniques

In our study, three classification techniques were used: k-nearest neighbors (k-NN), in general considered a benchmark classifier, and Support Vector Machines (SVM) and Multi-Layer Perceptron (MLP) neural network as state-of-the-art techniques. For an introduction to these techniques, we refer to the many handbooks and survey articles that have been published; we will only specify them here for purpose of replication.

9.3.1 k-Nearest Neighbors (k-NN)

We used WEKA's [252] k-NN implementation, based on Aha, Kibler, and Albert's instance-based learning algorithms [4]. In our study, its output was a probability of classification to each of the classes, but not the resulting class. In other words, if there was a single winning class, the output was 100% for the winning class and 0% for all the other classes. In the case of a tie between multiple classes, the output is divided between them and 0% is provided to the rest. All the recognition rates of the k-NN classifier reported here were obtained by using this modified algorithm.

A correct metric and an appropriate k are crucial parameters of a k-NN classifier. In the current study, the $1 - \mathrm{distance}$ distance weighting metric, a brute force neighbor search algorithm, and setting $k = 4$, provided the best results. For more information on k-NN, I refer to Appendix A.

9.3.2 Support vector machines (SVM)

One of its key parameters for SVM regularization is its cost parameter C (i.e., the cost of misclassifying points). This allows some flexibility in separating the classes as it determines the number of training errors permitted and, hence, it does or does not enforce rigorous margins. As such the parameter C determines the trade off between accuracy of the model on the training data and its ability to generalize. For this data set, C was set on 1.

Another key feature of SVM is its kernel function, which characterizes the shapes of possible subsets of inputs classified into one category [586]. Being SVM's similarity measure, the kernel function is the most important part of an SVM. We applied a radial basis function kernel, defined as:

$$k_G(x_i, x^l) = \exp\left(-\gamma |x_i - x^l|^2\right),$$

where x_i is a feature vector that has to be classified, x^l is a feature vector assigned to a class (i.e., the training sample), and γ is set to $1/28$, with 28 being the number of input parameters [586]. Note that the radial basis function is a variant of the Gaussian kernel function.

For the SVM, the LibSVM implementation [96] was used, using the cost parameter C and the kernel described here. For all other settings, the defaults of LibSVM were used [96]. For more information on SVM, I refer to Appendix A.

9.3.3 Multi-Layer Perceptron (MLP) neural network

We computed WEKA's [252] MLP trained by a back-propagation algorithm. It used gradient descent with moment and adaptive training parameters. For more information on artificial networks, MLP in particular, I refer to Appendix A. In our case, an MLP with 3 layers with 7 nodes in the hidden layer was shown to have optimal topology. This topology was trained with 500 cycles. For all other settings, the defaults of WEKA were used [252].

9.4 Results

Using the three classifiers introduced in the previous section, we conducted two series of analyses:

1. Cross-validation of the (precision of the) SUD with the parameters of the speech signal features that are classified by the k-NN, SVM, and MLP. On the one hand, this verifies the validity of the SUD; on the other hand, this determines the performance of the three classifiers in objective stress detection.
2. Classification of the happiness and fear conditions of both studies. This enables the inspection of the feasibility of CAD for PTSD. Additionally, analyses across both studies and of the baselines were conducted to inspect the effects of experimental design.

The input for the classifiers were the principal components described in the previous section. All classifiers were tested using 10-fold cross-validation, and their average performance is reported in Table 9.2.

9.4.1 Cross-validation

The SUD scale consisted of 11 bins (from 0 to 10). However, SUD score 10 was not used by any of the patients and, hence, could not be classified. So, for the classification 10 bins (i.e., SUD levels 0 to 9) were used. All three classifiers were successfully employed.

Assuming the SUD provides a valid comparison for the speech parameters, we classified the SUD scores over both studies, including both conditions and their baselines. All

Table 9.2: The classification results (in %) of k-nearest neighbors (k-NN), support vector machine (SVM) (see also Figure 9.4.1), and artificial neural network (ANN). Correct classification (C_N), baseline (or chance) level for classification (μ_N), and relative classification rate (C_N^*; see also Eq. 9.3) are reported. The Subjective Unit of Distress (SUD) was taken as ground truth, with several quantization schemes. N indicates the number of SUD levels.

N	μ_N	k-NN		SVM		ANN	
		C_N	C_N^*	C_N	C_N^*	C_N	C_N^*
2	50.00	89.74	79.74	89.74	79.47	82.37	64.74
3	33.33	74.74	124.21	78.16	134.47	72.37	117.11
4	25.00	68.42	173.68	66.32	165.26	57.37	129.47
5	20.00	53.42	167.11	55.00	175.00	48.95	144.74
6	16.67	52.63	215.79	53.42	220.53	47.63	185.79
7	14.29	44.74	213.16	47.11	229.74	42.37	196.58
8	12.50	42.89	243.16	43.16	245.26	41.58	232.63
9	11.11	42.89	286.05	44.21	297.89	34.74	212.63
10	10.00	38.95	289.47	38.68	286.84	36.32	263.16

classifiers had to be capable of detecting stress from speech, in particular when classification was simplified to the binary comparison of low versus high stress. The correct classification rate (C_N) by the k-NN, SVM, and MLP was, respectively, 89.74, 89.74, and 82.37 (see also Table 9.2).

Although the SUD is an established instrument in psychology, to our knowledge the precision of this instrument has not been assessed. The reliability of the SUD when aiming at a high precision of reporting, such as for a scale of 0-10, could be doubted if people's interoception is unreliable [132]. While this point is under debate [132], patients with anxiety disorders have recently been shown to be (over)sensitive to interoception [164].

In the current research, we not only used the SUD as a ground truth, but also quantized the scale into all possible numbers of levels, ranging from 10 to 2. This quantization is performed by discretizing the SUD responses into N steps, with a step size of r/N, where r is the range of the SUD values (i.e., 9). This quantization allows us to verify the reliability of the SUD in relation to the obtained speech parameters.

To provide a fair presentation of the classification results, we do not only provide the correct classification rate (C_N), but also the relative classification rate (C_N^*) for each of the N bins. The relative classification rate expresses the improvement of the classification compared to baseline (or chance) level. It is defined as:

$$C_N^* = \frac{C_N - \mu_N}{\mu_N} \times 100, \tag{9.3}$$

with μ_N being the baseline (or chance) level for N classes. This relative classification rate is also known as a range correction and used more often in health and emotion research [196].

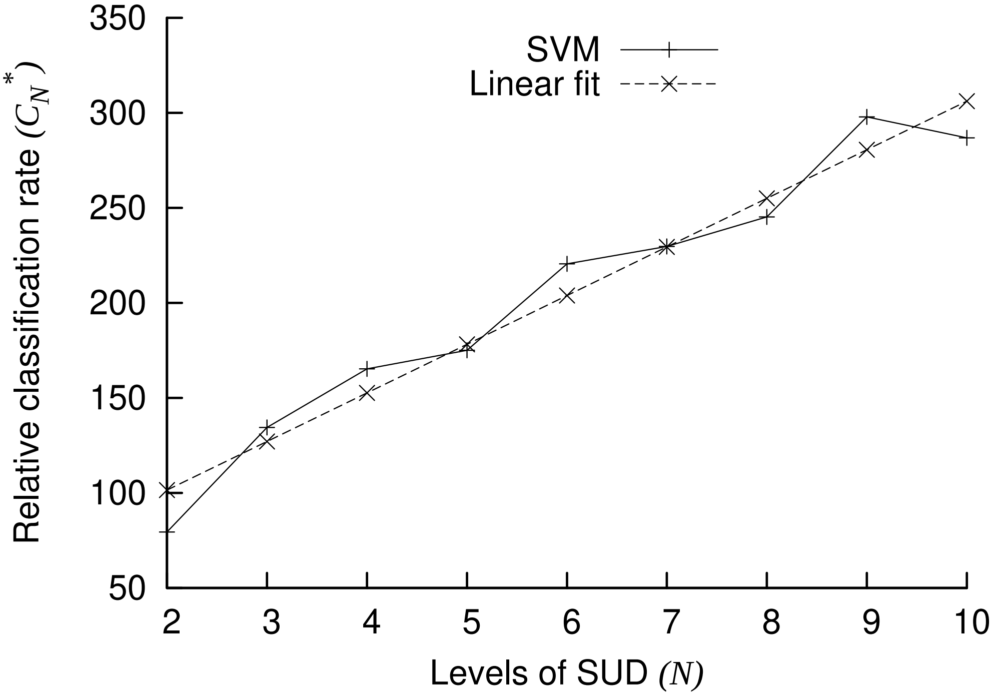


Figure 9.1: The overall relation between the reported Subjective Unit of Distress (SUD) and the relative correct classification using 11 principal components based on 28 parameters of speech features.

Consulting the relative classification rate (see Eq. 9.3) helps in determining the true classification performance on each level of quantization of the SUD as an assessor of the patient's distress level. The three classifiers show an almost monotone linear increase in relative classification rate; see Figure 9.4.1. The linear fit closely follows the data presented in Table 9.2 for all three classifiers (explained variance: $R^2 = .96$). This underlines the validity of the SUD as an instrument to assess people's stress levels. Moreover, it confirms its high concurrent validity, with its ability to discriminate between 10 levels of distress, and indicates that its use as ground truth for stress measurement is adequate.

9.4.2 Assessment of the experimental design

The two conditions of both studies in this paper functioned as triggers of stress and relaxation. The former study was meant to resemble a patient's behavior in one of his first therapy sessions; the latter the behavior of a patient in a late therapy session. The experimental design enabled us to conduct our research within a tight time window. This stands in sharp contrast with a longitudinal study, the only research alternative.

The success of the experimental design was assessed by classifying the PCA derived from the parameters of the speech signal features. All three classifiers (k-NN, SVM, and MLP) were applied. On the whole, the results of the MLP were disappointing compared to

Table 9.3: The classification results (in %) of k-nearest neighbors (k-NN) and support vector machine (SVM). Baseline (or chance) level for classification (μ_N), correct classification (C_N), and relative classification rate (C^*_N; see also Eq. 9.3) are reported. N takes either the value 2 or 3. Both the storytelling (ST) and reliving study (RL) analyzed, with $^+$ and $^-$ denoting respectively the happiness and stress triggering conditions.

	baseline	ST	ST$^+$	ST$^-$	RL	RL$^+$	RL$^-$	μ_N	C_N	C^*_N
k-NN	•	•						50.00	64.41	28.81
	•		•	•				33.33	41.95	25.85
			•	•				50.00	72.73	45.45
	•				•			50.00	84.58	68.17
	•					•	•	33.33	62.08	86.25
						•	•	50.00	72.92	45.83
SVM	•	•						50.00	64.83	29.66
	•		•	•				33.33	48.31	44.92
			•	•				50.00	88.64	77.27
	•				•			50.00	90.42	80.83
	•					•	•	33.33	59.58	78.75
						•	•	50.00	77.08	54.17

the k-NN and SVM and, as such, are of little value. Therefore, we will refrain from reporting the results for the MLP classifier and only report those for the k-NN and SVM classifiers. We separately compared the ST and the RL study with the baselines, which provided emotionally neutral speech signals.

A comparison between the two ST conditions and the baselines (taken together) revealed that they were very hard to distinguish (see also Table 9.3). This may be the case because the baselines consisted of reading a neutral story. Although ST has the advantage of a high level of experimental control, its disadvantage became evident as well: it had a limited ecological validity with respect to emotion elicitation. Classification of the ST conditions on the one hand, and of the baselines on the other, confirmed this finding with 64.41% (for the k-NN) and 64.83% (for the SVM) correct classification, respectively; see also Table 9.3. Classification of the two ST conditions only showed that these can be very well discriminated by the SVM: 88.64% correct classification, but less so by the k-NN: 72.73% correct classification; see also Table 9.3. These findings confirm that the neutral baseline ST laid between both ST conditions, as it was meant to be, but making it very hard to discriminate the three conditions.

Both RL conditions could be discriminated very well from the baselines (taken together) (see Table 9.3). Classification of the RL conditions on the one hand, and the baselines on the other, confirmed this finding with 84.58% (for the k-NN) and 90.42% (for the SVM) correct classification, respectively; see also Table 9.3. This result is in line with our expectations, because RL was shown to truly trigger emotions in patients suffering from PTSD. Although RL may allow less experimental control, its emotion-triggering turned out to be dominant. This finding stresses the need for ecologically valid research on mental health related issues. Classification results indicated also that it was harder to discriminate between the two RL conditions, by both the k-NN and SVM, with 72.92% and 77.08% correct classification, respectively; see also Table 9.3. In part, these results undermine the validity of the baselines for the reliving study, because other factors than emotion may have influenced the speech signal.

9.5 Discussion

We explored the feasibility of objective, ubiquitous stress assessment, which can help both in daily life and in therapy. To assure a controlled but ecologically valid assessment of stress, 25 PTSD patients participated in a controlled ST study and a RL study, each with a 'happy' and a 'stress triggering' session, as was exhaustively discussed in Chapter 8. The two sessions were meant to represent one of the first and one of the last therapy sessions a patient participates in. The stress level of the patients was assessed by two instruments: i) speech, as an objective and ubiquitous stress indicator and ii) the SUD, a clinically validated Likert scale. The SUD and speech model were cross-validated, using machine learning algorithms (i.e., k-NN, SVM, and MLP neural network). Correct classification rates of 90%, 78%, 44%, and 39% were achieved on, respectively, 2, 3, 9, and 10 SUD levels. Using the same classifiers, the two sessions could be discriminated in 89% (for ST) and 77% (for RL) of the cases. A clearer illustration of the difference in the level of complexity between (semi-)controlled and real-world studies could hardly be given.

The general validity of the two reported studies was high. Content validity of the studies was high, given that i) the studies aimed at a specific group of patients (i.e., PTSD), ii) the SUD and the speech signal features and their parameters were chosen with care (all were prominent in the literature), and iii) the cross-validation of the SUD with the speech signal features confirmed that they both provide a complete image of the patient's experienced stress. Criteria-related validity was also high, because speech was the preferred signal and can be recorded unobtrusively. The SUD scores were provided at a rate of one a minute, which can also be considered as accurate in the given context, as the stress level does not fluctuate that quickly. Ecological validity was maximized. For the RL study, we obtained natural stressors within a limited time window.

For decades, audio-based emotion recognition has been examined with a limited set of features-parameters (≤ 64) and without any feature selection or reduction [579, 696]. In the last decade, a brute force strategy using hundreds or even thousands of features (e.g., see [644, 725]) has been applied more often [590]. Together with the explosion in the number of features, feature selection/reduction strategies have claimed an increasingly important role.

A machine's recognition rate of emotional speech ranges from Banse and Scherer [27], who report $25\%/40\%$ correct classification on 14 emotions, to Wu, Falk, and Chan [725], who report $87\%/92\%$ correct classification on 7 emotions. The latter results, however, are in contrast with the results on a structured benchmark reported by Schuller, Batliner, Steidl, and Seppi [590] on the InterSpeech 2009 emotion challenge: $66\% - 71\%$ (2 classes) and $38\% - 44\%$ (5 classes). Apart from the differences in classification rate and the number of classes to be distinguished, these studies can both be questioned with respect to their ecological validity of the experienced emotions. In contrast, in at least one of our two studies (in particular, the RL study), true emotions were triggered. Furthermore, the ST study can be considered as half-way between common laboratory studies and real-world studies (like the RL study). Our classification results illustrated the considerable difference between the compromise ST study and the real-world RL study. They show that a careful interpretation of laboratory results is needed because a one-on-one mapping between lab and real-world results cannot be taken for granted.

An alternative explanation for the differences between the ST and RL studies can be sought in the expression of emotions rather than in their experience. Already in 1908, Anton Marty [439] proposed a differentiation between emotional and emotive communication. In emotional communication, speech serves as a spontaneous, unintentional leakage or bursting out of emotion. In contrast, in emotive communication speech there is no automatic or necessary relation to "real" inner affective states. As such, emotive communication is considered to be a strategy to signal affective information in speech. It uses signal patterns that differ significantly from spontaneous, emotional expressions, which can be initiated both intentionally and unintentionally [27, 334]. Possibly, emotional communication was dominant in the RL study and emotive communication in the ST study. Further research may reveal whether this distinction underlies the differences in classification in the two studies that we observed.

9.6 Conclusion

In this paper, we have presented two studies involving one and the same group of PTSD patients. This experimental design provided us with two unique but comparable data sets that only differed with respect to task. As such, a comparison of two stress elicitation methods, ST and RL, was possible. The comparison revealed both commonalities and differences between the two studies, which are directly relevant to several theoretical frameworks, such as the ones outlined just before in the discussion.

It would be of interest to apply the models developed in this research to patients suffering from other related psychiatric disorders, such as depression [9, 333, 483], insomnia [9], and generalized anxiety disorder [9, 506]. Probably, even for less related psychiatric disorders, the current approach would be a good starting point. In such a case, the general framework and speech signal processing scheme, as presented in this paper, could be employed. Most likely, only the set of parameters used for the processing pipeline would have to be tailored to the specific disorders.

Apart from being unobtrusive, the speech signal processing approach, as applied in the current studies, has another major advantage: it enables the remote determination of people's stress. This feature enables its use in yet another range of contexts; for instance, in telepsychiatry [279, 483], as personal stress indicator [17, 483], and in call-centers [182, 463] that frequently have to cope with highly agitated customers. However, as for the different psychiatric disorders and the other application areas mentioned, the processing pipeline should be adapted to this situation as well.

Taken together, an important and significant step was made towards modeling stress through an acoustic model, which can be applied in our daily lives and in mental health care settings. By the specific research design, it was ensured that "real" stress was measured. In addition, both precise subjective measurement using the SUD, as well as objective measurement through speech signal processing, were shown to be feasible to detect stress and as such determine therapy progress in an unobtrusive manner. Statistical models were constructed on the basis of a selection from 78 parameters of five speech features, which showed reliable and robust stress classification. In sum, we hope to have shown that unobtrusive and ubiquitous automatic assessment of emotion and experienced stress is possible and promising.

With the next part of this monograph, Part V, I will close this monograph. This part consists of two chapters. In the next chapter, Chapter 10, I propose a set of general guidelines for *ASP* and *affective computing* in general. These guidelines will stress, in separate sections, signal processing and pattern recognition issues. The second and last chapter of Part V, Chapter 11, is a general discussion with which I close this monograph.

V. EPILOGUE

10

Guidelines for ASP

Abstract

Although emotions are embraced by science, their recognition has not yet reached a satisfying level. Through a concise overview of affect followed by a set of studies, we provided some insight into the problems encountered. In this chapter, we will identify guidelines for successful *Affective Signal Processing (ASP)*. First we will discuss: physical sensing characteristics, temporal construction, normalization, and context. Second and last, guidelines for successful classification of emotions will be presented, which will include validation (e.g., mapping of constructs on signals), triangulation, and user identification. With this concise set of directives for future research in *affective computing*, I will present important conclusions drawn on my experiences throughout almost a decade of research. I hope that these guidelines may help in the further maturation of the field.

This chapter is an extended version of the fifth and sixth section of:

Broek, E. L. van den, Janssen, J.H., Zwaag, M.D. van der, Westerink, J.H.D.M., & Healey, J.A. Affective Signal Processing (ASP): A user manual. *[in preparation]*

which already appeared partially as:

Broek, E.L. van den *et al.* (2009/2010/2011). Prerequisites for Affective Signal Processing (ASP) - Parts I-V. In A. Fred, J. Filipe, and H. Gamboa, *Proceedings of BioSTEC 2009/2010/2011: Proceedings of the International Joint Conference on Biomedical Engineering Systems and Technologies*. January, Porto, Portugal / Valencia, Spain / Rome, Italy.

10.1 Introduction

This monograph started with an introduction on *ASP*, stressed its importance for three branches of computer science, and introduced a closed loop model as working model for emotion-aware systems. Chapter 2 reviewed *affective computing* in its broadest sense and concluded: i) *affective computing* lacks the required progress and ii) *ASP* is a promising (covert) channel for *affective computing*. This initiated a quest for *ASP*'s pitfalls and the reasons for its lack of progress. In the six chapters that followed, Chapters 3-7, studies were presented that explored various aspects of *ASP*. Although this vast amount of work is by no means an exhaustive investigation of all aspects of ASP, it did reveal a range of important results. These results have been taken together and their implications beyond those of the specific studies have been considered. This has resulted in a set of guidelines for *ASP*, which will be introduced in this chapter.

This chapter will consist of two main sections, which will address the two components of the signal processing + pattern recognition pipeline that forms the core of closed loop systems, as was denoted in Chapter 1 (see Figure 1.1). For both components, guidelines will be provided with the aim to bring *ASP* from research to practice. The chapter will close with a brief conclusion.

10.2 Signal processing guidelines

Signal processing is the first essential phase of the signal processing + pattern recognition pipeline, as was already denoted in Section 1.5 (see also Figure 1.2). This section identifies four issues that should be considered in *Affective Signal Processing (ASP)*, namely: physical sensing characteristics, temporal construction (including: intertwined psychological and physiological processes and time windows), normalization, and context. If the guidelines are met, high quality source signals can be recorded and proper feature extraction can be applied.

10.2.1 Physical sensing characteristics

In this section, we will discuss the implications of the physical sensing characteristics of sensors and the environment for *ASP*. There are a number of different sensors [103]. For respiration measurements, a gauge band can be placed around the chest. Thermistor sensors placed on the surface of the skin can be used to measure skin temperature [346]. HR can be measured through surface electrodes (ECG) or through a photoplethysmography. Skin conductance and muscle tension (EMG) are also measured through surface electrodes.

The choice of surface electrodes depends on the kind of measurement, the aim of the measurement, and the application in which it is to be used (e.g., see [666]). For example, in the lab one opts for the most sensitive and reliable electrodes, which are wet electrodes that use a gel for better conductivity. However, for wearable affective measurements dry electrodes are a better option, as these are more practical and easier to attach and incorporate into devices. The kind of gel used with wet electrodes depends on the measurement type. For skin conductance measurements, a saltless gel should be used as salt changes the composition of the skin which influences the measurement [62, 234]. For EMG and ECG, gels with high electrical conductance are better, hence they often include salt.

The location of the surface electrodes is important as improper placing can cause noise in the signal [85, 294, 376, 449] but the size of the electrodes can also be of influence [456]. However, in the case of *ASP*, the wearable devices and setting will put constraints on the location and size of the sensors. For example, the upper phalanx of the finger tips conventionally used for skin conductance measurements cannot be used while driving a car [269, 272, 329, 330, 474]. Other parts of the hands or the sole of the foot might be used instead [62, 163]. However, the number of sweat glands differs significantly between these and other possible positions of measurement; see also Table 10.1. Skin temperature can also be measured on the foot instead of the hand. However, similar as with EDA measurements,

Table 10.1: Distribution of eccrine (sweat) glands in man, adopted from [561, Chapter 6].

location	# glands	Location	# glands
palms	2,736	dorsa of the feet	924
soles	2,685	thigh and leg, medical aspect	576
dorsa of hands	1,490	thigh, lateral aspect	554
forehead	1,258	cheek	548
chest and abdomen	1,136	nape of neck	417
forearm, flexor aspect	1,123	back and buttocks	417
forearm, extensor aspect	1,093		

Table 10.2: Results of a representative study on the influence of climate on the number of sweat glands, adopted from [561, Chapter 6].

subjects	# subjects	age	# sweat glands ($\times 1,000$)		
			min	max	mean
Japanese in Japan	12	13–16	1,069	1,991	1,443 $\pm$ 52
Japanese migrated to the tropics	6	38–58	1,639	2,137	1,886 $\pm$ 59
Japanese born in tropics	4	14–35	1,839	2,603	2,168 $\pm$ 11
Ainu*	11	6–35	1,781	2,756	2,282 $\pm$ 66
Russians	10	17–42	2,642	2,982	2,800 $\pm$ 23
Filipino	15	9–25	2,589	4,026	2,961 $\pm$ 61

* The Ainu is a Japanese aboriginal tribe living in the cold northern parts of Japan.

the location is of influence on the measurements [561, Chapters 10]; see also Table 10.3. The thermal circulation index (CI), as provided in Table 10.3, can be used to compare measurements done on distinct locations. For HR, oxygen saturation, and HRV, instead of using electrodes on the chest (ECG) one can use a photoplethysmographic sensor on the ear, hand, or foot [19]. The last decade several photoplethysmographic sensors have been introduced that have shown to be reliable, where this was not yet the case with photoplethysmographic sensors introduced in the previous century [19].

The physical sensing characteristics of the environment such as humidity and temperature also play an important role. This predominantly influences the skin conductance and temperature measurements [561, Chapters 6 and 10]. Moreover, people's number of sweat glands depends heavily on the climate people are living in; see Table 10.2. These points are of special interest for longer periods of continuous measurement and is also different in medical experiments, which require a controlled lab situation in which the humidity and temperature of the room can be kept constant. Hence, it might be worthwhile to enhance unobtrusive sensor platforms with environmental sensors. Another way of dealing with the issues of different sensor positions and changes in environmental temperature and humidity can be to standardize the measurements using z-scores for each session. During continuous longer term measurements one can use a sliding time window for a set period (e.g., one or two hours), which is used for standardization.

To conclude, due to the great number of differences in the aim of physiological measurements, different sensor positions, and different or even changing environmental conditions, one should always carefully puzzle to find the best combination of electrode types

Table 10.3: Results of a representative study on skin temperature (in ^{o}C) and thermal circulation index (CI) (i.e., CI = Δ(skin,air) / Δ(interior,skin)) in relation to several body regions, adopted from [561, Chapter 10] Room temperature was 22.8 ^{o}C and rectal temperature (as reference temperature) was 37.25^{o}C.

body region	temperature (in ^{o}C)			CI
	skin	Δ(skin,air)	Δ(interior,skin)	
forehead	33.40	10.60	3.85	2.75
clavicle	33.60	10.80	3.65	2.96
over breast	32.75	9.95	4.50	2.21
1 inch over umbilicus	34.20	11.40	3.05	3.75
over apex of heart	33.30	10.50	3.95	2.67
lumbar region	33.30	10.50	3.95	2.67
arm, biceps	32.85	10.05	4.40	2.28
palm of the hand	32.85	10.05	4.40	2.28
kneecap	32.35	9.55	4.90	1.95
calf of leg	32.20	9.40	5.05	1.86
sole of foot	30.20	7.40	7.05	1.05
big toe	30.95	8.15	6.30	1.29

Table 10.4: Eight methods to normalize affective signals. x denotes the (original) signal and $\min$ and $\max$ are its (estimated) minimum and the maximum. μ_B, min_B, max_B, and σ_B are respectively the mean, minimum, maximum, and standard deviation of the baseline.

1	$\log x_i + 1$ or $\sqrt{x_i}$	To correct for positively skewed and leptokurtotic signals.
2	$x_i - \mu_B$	A standard correction, known as delta or reaction scores.
3	$x_i - \min$	A useful alternative to the first method, when there is no baseline measurement (and a lot of variance in the signal).
4	$x_i - \mu_B/\sigma_B$	Known as standardization or the z-correction. It significantly reduces the variance in the signal. Also used as range correction.
5	$x_i - \mu_B/\mu_B$	Standard range correction.
6	$x_i - \min / \max - \min$	Correction for individual differences in physiology. Sensitive to outliers.
7	$x_i / \max$	Used for Skin conductance responses features. In practice, the *max* is determined by frightening a subject at the start.
8	$(x_i - \mu_B/\mu_B) \times 100$	Percentage change, used as range correction.

Sources of information: 1: [85, Chapter 7; p. 165], 2: [418], 3 and 7: [62], 4 [62, 77], 6: [430, 449], and 8: [196] (see also [206]).

and locations. Furthermore, standardizing the signals will also reduce a lot of the otherwise unexplained variance in the signal; for example, normalization techniques, see Table 10.4. In the end, this will provide cleaner signals to the machine learning algorithms and will lead to a much more successful *ASP*.

10.2.2 Temporal construction

There are many temporal aspects in *ASP* that should be taken into account, as will be discussed here. In order of appearance, we will discuss: habituation, Law of Initial Values (LIV), sources of delay in measurement, and time window selection [585, Chapter 23].

Humans are not linear time (translation or shift) invariant systems [62], they have a tendency for habituation [26, 198]. This increases the complexity of *ASP* substantially, since most signal processing techniques rely on that assumption. In general, every time a stimulus is perceived one's reaction to it will decrease. With large delays between the stimuli, one recovers from the habituation effect. When it becomes possible to track emotional events, this information can be used to predict how strong the effect of a similar stimulus will be.

Physiological activity tends to move to a stable neutral state. This principle is known as the LIV [721]; see also [626, Chapter 5] and [144, Chapter 12]. For example, when you perceive a scary stimulus when your heart rate is at 80 it might increase by 15 beats per minute; in contrast, if your heart rate is at 160 it is unlikely to increase at all. So, the effect of a stimulus on physiology depends on the physiological level before stimulus onset, which has been shown to be a linear relationship that can be modeled by linear regression: determine a regression line (which is different per feature and person) and, subsequently, use it to correct each value by computing its residualized value [316, 546, 721].

There are many challenges in modeling the temporal aspects of emotion, among which the following triplet:

1. Affective signals are typically derived through non-invasive methods to determine changes in physiology (cf. [212, 213]) and, as such, are indirect measures. Hence, a delay between the actual change in emotional state and the recorded change in signal has to be taken into account.
2. The annotation challenge:
   (a) How to determine when an emotion begins and when it ends? and
   (b) A delay in reporting the feeling by the subject, an issue of postdictive validity (see Section 10.3.1). In practice, time window selection can be done empirically, either manually or automatically; for example, by finding the nearest significant local minima or making assumptions about the start time and duration of the emotion. A solution would be to ask subjects themselves to define the window of interest.

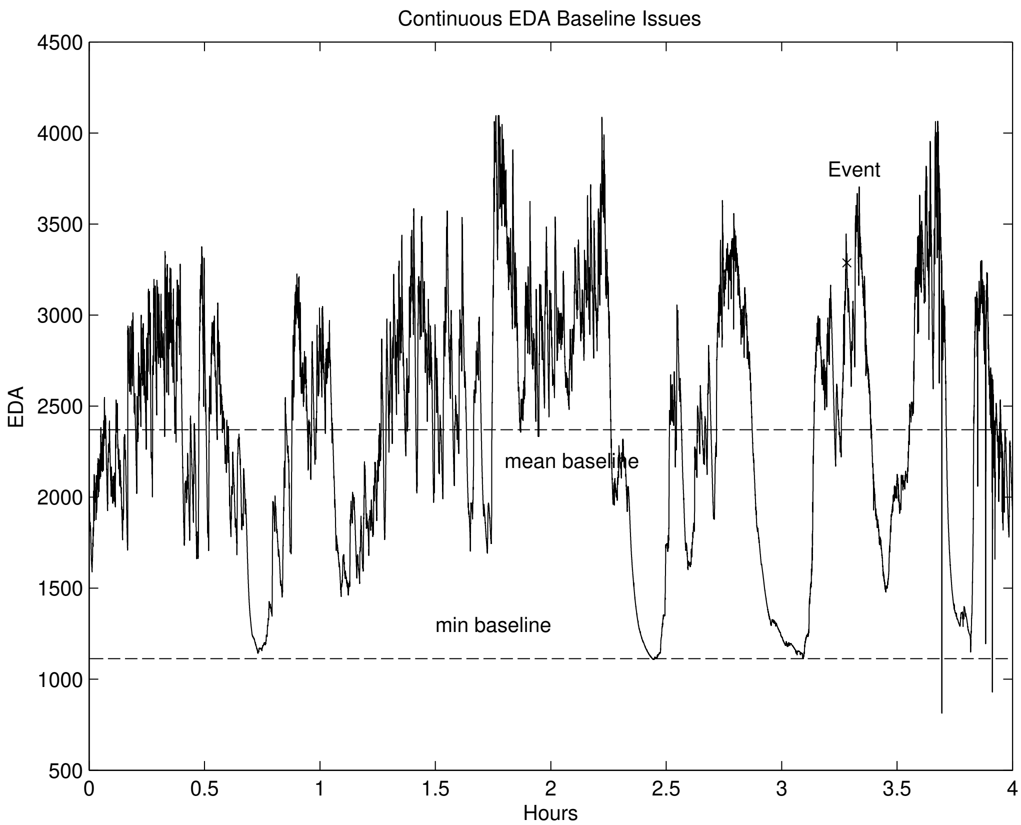


Figure 10.1: Four hours ambulatory EDA recordings, with its minimum and mean baseline.

3. The so-called sensor fusion problem: one has to determine how to window individual signals within the emotional event since different signals have different latencies; see also Table 1.1.

It should be noted that, in general, biosignal features that are calculated over time windows with different lengths cannot be compared with each other (e.g., see Figures 10.1–10.3). Therefore, it is important to keep the window length constant, while processing an event.

Considerations that need to be taken into account when selecting time windows include the following three:

1. The psychological construct studied, as different psychological processes develop over different time scales. On the one hand, emotions lead to very short and fast phasic changes and, thus, require short time windows (i.e., in the order of seconds; cf. Figures 10.1–10.3). On the other hand, changes in mood are more gradual and tonic and, so, require broader time windows (e.g., hours or at least minutes; see Figures 10.1–10.4). Moreover, some changes can hold for only a brief moment, while others can even be permanent.

2. The type of signal being measured; see also Table 1.1. At least the minimum response time of the signal should be taken as time window but preferably longer. In general, the longer the duration interval, the more reliable the data becomes; and

3. The context (see also Section 10.2.4). Discontinuous signals (e.g., heart rate variability) require longer time windows than continuous signals (e.g., skin conductance and skin temperature). In ambulatory settings, this distinction becomes even more important as discontinuous signals are prone to disturbances and, hence, shorter time windows are unreliable (cf. 1 and 2 minutes for HF and LF in controlled and 5 minutes in ambulatory settings is advised [641]), where this issue is less pregnant for continuous signals; see also Table 1.1.

The results shown in Table 10.5 that accompany Figures 10.2 and 10.3 illustrate the significant impact the choice of time window length can have on calculating features, especially those related to signal shape. This illustrations shows the impact time windows can have on *ASP* and, hence, the importance of taking into account the considerations presented here.

Statistic	Time windows (in seconds)		
	5	10	60
mean	3314	3312	3083
SD	19	23	217
slope	43	-69	697

Table 10.5: Standard statistics on three time windows of an EDA signal, as presented in Figure 10.3. These three time windows are close-ups of the signal presented in Figure 10.2, which in turn is a fragment of the signal presented in Figure 10.1.
Note. SD denotes Standard Deviation.

10.2.3 Normalization

Finding an appropriate normalization method is both important and difficult for sensors whose readings depend on factors that can easily change on a daily basis, such as sensor placement, humidity, temperature, and the use of contact gel, as was already noted in Section 10.2.1. Physiological signals can be normalized using:

- Baseline corrections: applied when comparing or generalizing multiple measurements from one individual across a variety of tasks [418].
- Range corrections: reduce the inter individual variance by a transformation that sets each signal value to a proportion of the intra individual range [62].

Probably the most frequently used and powerful correction for continuous biosignals (e.g., EDA and skin temperature) is standardization (method 4 in Table 10.4) [62]. It corrects not only for the baseline level but also for the variation in the signal, making it robust.

Other correction methods are tailored to specific features; for example, the amplitude of skin conductance responses is often corrected by dividing by the maximum amplitude. An alternative is the use of delta, or reaction scores, (Table 10.4, nr. 2), which is suitable

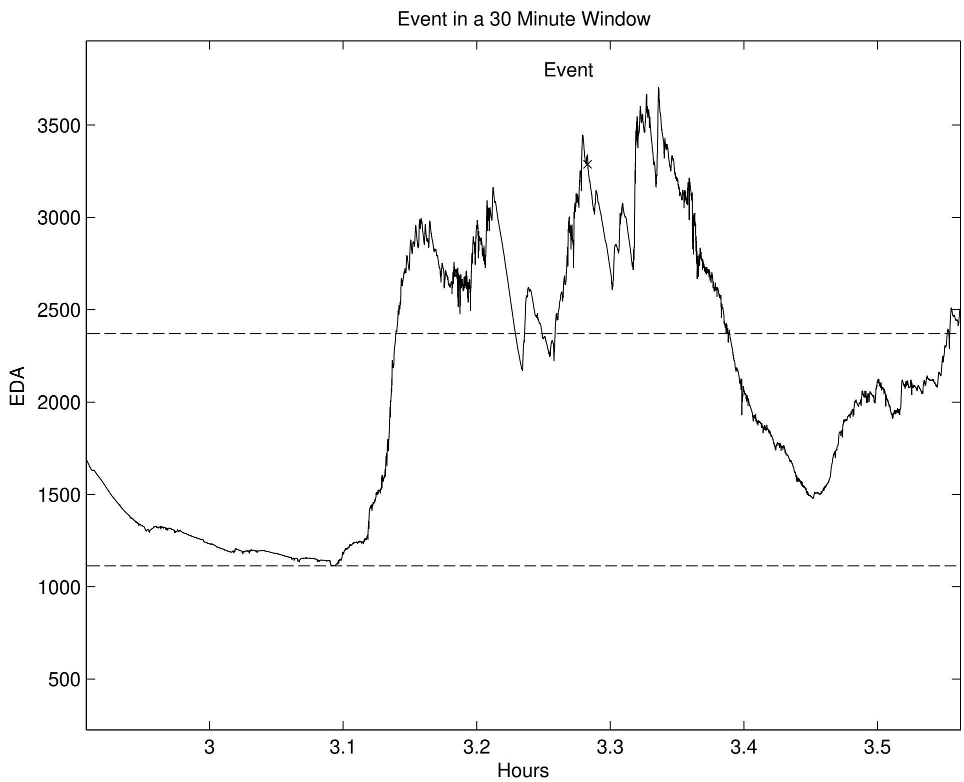


Figure 10.2: A 30 minute time window of an EDA signal, which is a part near the end of the signal presented in Figure 10.1. Three close-ups around the event near 3.3 hours are presented in Figure 10.3.

and reliable for absolute level comparisons [418]. If no baseline measurements are available, method nr. 3 of Table 10.4 is a good alternative.

Normalization methods $4 - 8$ in Table 10.4 are often used as range corrections. In general, they provide a stronger normalization than baseline corrections 2 and 3 in Table 10.4. As such, range corrections can also be used to compensate for greater variability in signals. Typical measures that are subject to large inter individual differences are skin conductance (tonic levels vary per person: $2 - 16\mu S$), skin temperature, and pulse volume.

Selecting a normalization method is difficult since each has different merits; see Table 10.4. Taking the minimum baseline is more equivalent to taking the resting EDA that would normally be used in a laboratory experiment. This is the best method if a consistent minimum seems apparent in all data being combined. The problem is that for each data segment, a minimum must be apparent. It is straightforward to eliminate point outliers such as those at 3.7 hours and 3.9 hours in Figure 10.2 and find a more robust minimum baseline.

With choosing an appropriate normalization method, the selection of a period (i.e., a time window) over which to calculate the parameters of the selection method (the normalization period) is also of importance, as was depicted in Section 10.2.2. As an example, Figures 10.2 and 10.4 show several hours of an ambulatory EDA signal, along with two

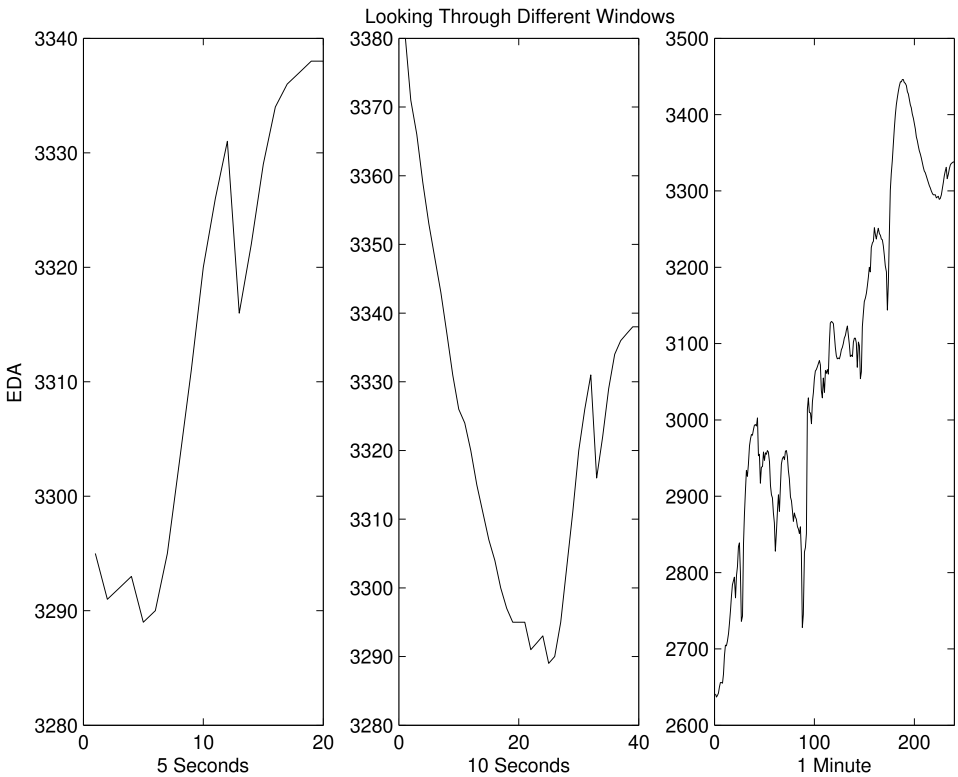


Figure 10.3: Three close-ups around the event presented in Figure 10.2. The statistics accompanying the three close-ups can be found in Table 10.5.

strategies for baseline correction: a minimum value baseline and a mean baseline (methods 2 and 3 in Table 10.4). Once the baseline is removed, the signal becomes a new base (or zero) and the original value is lost.

For short term experiments, a single baseline period is usually sufficient. However, when monitoring continuously, the baseline may have to be re-evaluated with greater frequency. The challenge here is to find a good strategy for dividing the signal into segments over which the baseline should be re-calculated. A simple solution is to use a sliding window; for example, where the last 30 minutes are taken into account. However, Figure 10.4 shows an obvious problem with this: the problem of lost data (e.g., sensor which has fallen off); see also Section 10.2.2. In sum, at this moment the most useful correction methods for each individual physiological measurement should still be specified. The most useful technique depends on the aim of the study.

10.2.4 Context

"*When humans talk with humans, they are able to use implicit situational information, or* context, *to increase the conversational bandwidth. Unfortunately, this ability to convey ideas does not transfer well to humans interacting with computers. In traditional interactive computing, users have an impoverished mechanism for providing input to computers. Consequently, computers are not currently enabled to take full advantage of the context of the human-computer dialogue. By improving the computer's access to context, we increase the richness of communication in human-computer interaction and make it possible to produce more useful computational services.*" A. K. Dey [158, p. 4] If anything, the experience and transmission of emotions via biosignals depends heavily on context [585, Chapter 23]. However, as is stated in the quote above, capturing context is easier said than done [6, 325, 668, 669]. Handling context is even considered to be one of AI's traditional struggles [649, 675]. Perhaps this can be attributed partly to the fact that in the vast majority of cases, research on context aware computing has taken a technology-centered perspective as opposed to a human-centered perspective [383]. This technology push has been fruitful though, among many other techniques, sensor networks, body area networks, GPS, and RFID have been developed. Their use can be considered as a first step towards context aware computing. However, not only is the gathering challenging but processing (e.g., feature extraction) and interpretation are also hard [21, 676, 699].

Potentially, context aware computing can aid *ASP* significantly. Biosensors can be embedded in jewelery (e.g., a ring or necklace), in consumer electronics (e.g., a cell phone or music player), or otherwise as wearables (e.g., embedded in cloths or as part of a body area network). Connected to (more powerful) processing units they can record, tag, and interpret events [158] and, in parallel, tap into our emotional reactions through our physiological responses. However, affective biosignals are influenced by (the interaction between) a variety

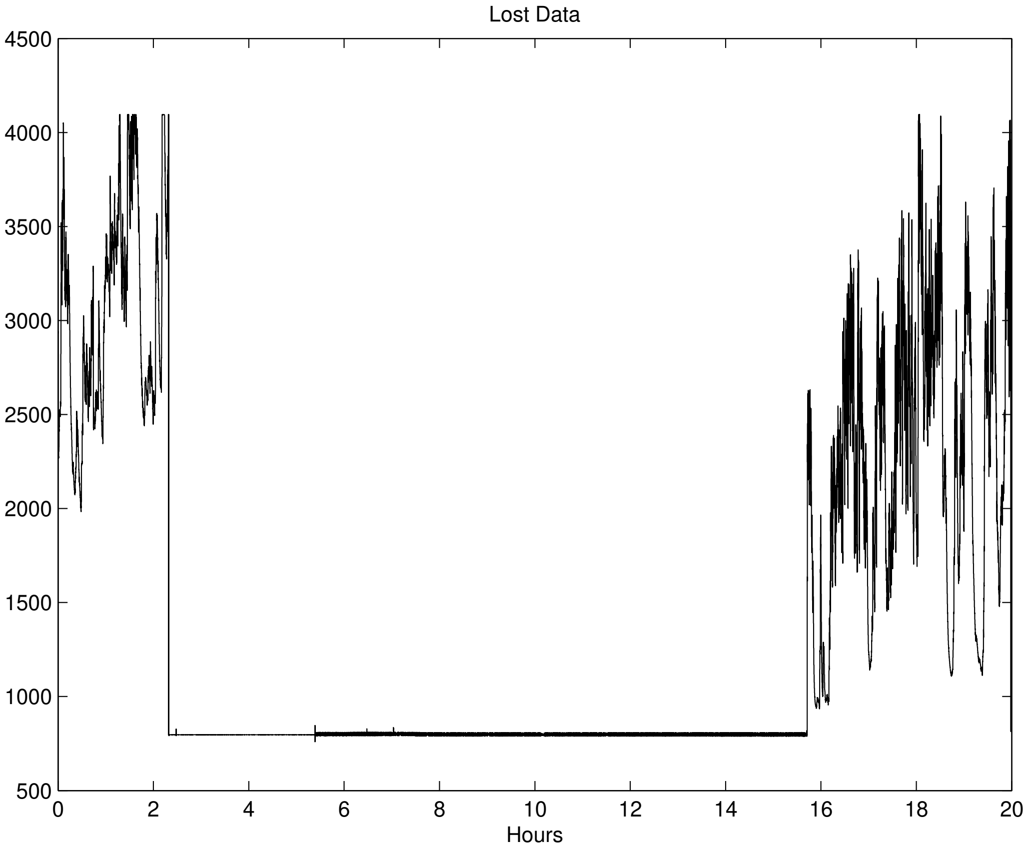


Figure 10.4: A typical sample of lost data with an EDA signal, as frequently occurs in real-world recordings.

of factors besides affect [78], as has been illustrated throughout this monograph.

To bring the theory just presented into practice, we present an example on the level of activity, as a factor of context. Figure 2.1 illustrates how pervasive motion artifacts can be for *ASP* in real world settings. Both heart rate and electrodermal activity are elevated during the period of high activity (i.e., from 27 to 30 minutes), as automatically determined through accelerometers. However, as the signal graphs show, the changes in heart rate follow changes in activity much more rapidly than electrodermal activity does, both in terms of onset and, especially, in terms of recovery. For level 4 (walking) in Figure 2.1, it even seems that the physical effects are so dominant that *ASP* should not be attempted. In contrast, with levels 1 (lying down), 2 (sitting), and 3 (standing/strolling) this is possible. So, physical activity can easily cast a shadow over affective (bio)signals.

10.3 Pattern recognition guidelines

As is illustrated in Table 2.4, a plethora of feature selection algorithms and classifiers has been applied in *affective computing* studies. Much has been said on the pros and cons of the methods applied, each article is accompanied with its own reasoning on it or simply ignores possible alternatives. For example, Picard and colleagues [268, 271, 524, 703] posed that sequential floating forward search (SFFS) was superior to 'standard' stepwise feature selection (SFS). This claim was questioned in follow-up research by others and recently rejected by Way, Sahiner, Hadjiiski, and Chan [708], who concluded: *"PCA was comparable to or better than SFS and SFFS for LDA at small samples sizes"* (p. 907) and *"In general, the SFFS method was comparable to the SFS method ..."* (p. 907). However, it should be noted that Way et al. [708] concerned a(n excellent) simulation study not in the domain of *affective computing*. Nevertheless, it illustrates the still ongoing debate on pattern recognition and machine learning methods. In the choice of these methods, pragmatic considerations, personal preferences, science's current fashion often seem to be dominant factors of choice. Moreover, several excellent handbooks are available on pattern recognition and machine learning [48, 170, 457, 648, 689] as well as a range of excellent tutorial and survey articles. Therefore, we will refrain from providing an overview of these techniques and provide general but crucial guidelines for the classification of affective signals, which are often violated. We pose that the triplet of guidelines can significantly improve *ASP*.

10.3.1 Validation

In the pursuit of a method to trigger emotions in a more or less controlled manner, a range of methods have been applied: actors, images (IAPS) (see Chapter 5), sounds (e.g., music) [316, 681], (fragments of) movies (see Chapters 3, 4, 6, and 7), speech [677], commercials [529], games (including serious gaming), agents, virtual reality [86, 474, 488, 616], reliving of emotions (see Chapters 8 and 9), and real world experiences [269, 270, 272, 316]; see also Table 2.4. However, how can we know which of these methods actually triggered participants' true emotions? This is a typical concern of validity, which is a crucial issue for *ASP*. For *ASP* purposes, validity can best be obtained through four approaches: content, criteria-related, construct, and ecological validation, which I will discuss next.

Content validity refers to a) The agreement of experts on the domain of interest (e.g., limited to a specific application or group of people, such as twins [427–429]); b) The degree to which a feature (or its parameters) of a given signal represents a construct; and c) The degree to which a set of features (or their parameters) of a given set of signals adequately represents all facets of the domain. For instance, employing only skin conductance level (SCL) for *ASP* will lead to a weak content validity when trying to measure emotion, as SCL is known to relate to the arousal component of an emotion, but not to the valence component. However, when trying to measure only emotional arousal, measuring only SCL may form strong content validity.

Criteria-related validity handles the quality of the translation from the preferred measurement (e.g., ECG) to an alternative (e.g., BVP), rather than to what extent the measurement represents a construct (e.g., a dimension of emotion space). Emotions are preferably measured at the moment they occur; however, measurements before (predictive) or after (postdictive) the particular event are sometimes more feasible (e.g., through subjective questionnaires). The quality of these translations are referred to as respectively predictive or postdictive validity. A third form of criteria-related validity is concurrent validity: a metric for the reliability of measurements (e.g., EDA recording on the foot sole) applied in relation to the preferred standard (e.g., EDA recording on the hand palm). For instance, the more affective states are discriminated the higher the concurrent validity.

A construct validation process aims to develop a nomological network (i.e., a ground truth) or an ontology or semantic network, built around the construct of interest. Such a network requires theoretically grounded, observable, operational definitions of all constructs and the relations between them. Such a network aims to provide a verifiable theoretical framework. The lack of such a network is one of the most pregnant problems *ASP* is coping with. This problem has been assessed in Chapters 5 and 6 that applied both the valence-arousal model and basic emotion categories as representations for affective states. A frequently occurring mistake is that emotions are denoted, where moods (i.e., longer object-unrelated affective states with very different physiology) are meant. This is very relevant for *ASP*, as it is known that moods are accompanied by very different physiological patterns than emotions are [223].

Ecological validity refers to the influence of the context on measurements. We identify two issues: 1) Natural affective events are sparse, which makes it hard to let participants cycle through a range of affective states in a limited time frame; and 2) The affective signals that occur are easily contaminated by contextual factors; so, using a context similar to that of the intended *ASP* application for initial learning is of vital importance. Although understandable from a measurement-feasibility perspective, emotion measurements are often taken in controlled laboratory settings. This makes results poorly generalizable to real-world applications.

The concern of validity touches upon the essence of research. However, it is still frequently ignored in branches of computer science. With this guideline, I hope to have provided some workable definitions of four types of validity that are crucial for *ASP*. These four types of validity should be respected, both when conducting research and when developing applications.

10.3.2 Triangulation

Triangulation is the strategy of combining multiple data sources, investigators, methodological approaches, theoretical perspectives, or analytical methods within the same study

[344, 651]. This provides the methodological instruments to "*separate the construct under consideration from other irrelevancies in the operationalization*" [273, p. 15901]. We propose to adopt this principle of triangulation, as applied in the social sciences and human-computer interaction, for *ASP*. Within the domain of *affective computing*, the constructs under investigation are emotions and irrelevancies can be the various sources of noise, as were mentioned in Chapter 2 and in Section 10.2.1.

Generally, five types of triangulation are distinguished [344, 651], each having their own advantages, namely:

1. *Data triangulation*: Three dimensions can be distinguished in data sources: time, space (or setting), and person (i.e., the one who obtained the recordings) [154]. Time triangulation can be applied when data is collected at different times [344]; for example, as is done by Picard et al. [524], Healey and Picard [272], and Janssen, Van den Broek, and Westerink [316]. In general, variance in events, situations, times, places, and persons are considered as sources of noise; however, they can also add to the study. Extrapolations on multiple data sets can provide more certainty in such cases. In turn, corrections can also be made to atypical data in a result set that clearly deviates from other results [651].

2. *Investigator triangulation*: Multiple observers, interviewers, coders, or data analysts can participate in the study. Agreement between these researchers, without prior discussion or collaboration with one another, increases the credibility of the observations [154]. Par excellence, this type of triangulation can be employed on including context and unveiling events as this often includes subjective interpretations of events, see also Section 10.2.4.

3. *Methodological triangulation*: This can refer to either data collection methods or research designs [404]. The major advantage is that deficiencies and biases that stem from a single method can be countered [651]. Multiple data sets (e.g., both qualitative and quantitative) and signal processing techniques (e.g., in the time and spectral domain) can be employed (see Table 2.4). Moreover, multiple feature extraction paradigms, feature reduction algorithms, and classification schemes can be employed (again, see Table 2.4). Further, note that methodological triangulation is also called multi-method, mixed-method, and methods triangulation [233].

4. *Theoretical triangulation*: Employing multiple theoretical frameworks when examining a phenomenon [154, 301, 396]; for example, using both a categorical (or discrete) and a continuous (e.g., valance-arousal) model of emotion [673, 676]. See Chapters 5 and 6 for a discussion on this topic.

5. *Analytical triangulation*: The combination of multiple methods or classification methods to analyze data [344, 682]. As is shown in Table 2.4, this approach has already

often been employed. For example, Picard et al. [524] and Healey and Picard [272] combined different signals from the same modality and Kapoor, Burleson, and Picard [328] and Bailenson et al. [25] combined biosignals with a vision-based approach. Paulmann, Titone, and Pell [512] who combined speech processing and eye tracking, which revealed that emotional prosody has a rapid impact on gaze behavior during social information processing. This facilitates (cross) validation of data sources, as is also described in Section 10.3.1.

In general, in well controlled research, we advise the recording of at least 3 affective biosignals and the derivation of at least 3 features from them, for each construct under investigation. In ambulatory, real-world research much more noise will be recorded, as also described in Chapter 2 and Section 10.2.1. To ensure that this noise can be canceled out, we advise the recording of many more affective biosignals and also the extraction of more features from them. As a rule of thumb for ambulatory research we advise researchers to record as many signals possible, avoiding interference with participants' natural behavior. However, a disadvantage accompanies this advice, as "*a 'more is better' mentality may result in diluting the possible effectiveness of triangulation*" [651, p. 256] Moreover, where possible, qualitative and subjective measures should always accompany the signals (e.g., questionnaires, video recordings, interviews, and Likert scales); for example, see [272, 616, 677, 716].

10.3.3 User identification

Throughout the field of *affective computing*, there is a considerable debate on present on generic versus personal approaches to emotion recognition. Some research groups specialized in *affective computing* have moved from general *affective computing* to *affective computing* for specialized groups or individuals. In general, the identification of users has major implications for *ASP*. We propose three distinct categories, from which research in affective science could choose:

1. *all*: generic *ASP*; see also Table 2.4 and [676, 679, 681]
2. *group*: tailored *ASP*; for example, see [104, 188, 274, 354, 592, 627, 633]
3. *individual*: personalized *ASP*; for example, see [40, 272, 316, 427–430, 464, 524, 624]

Although attractive from a practical point of view, the category *all* will probably not solve the mysteries concerning affect. It has long been known in physiology, neurology, and psychology that special cases can help in improving *ASP* [633]. For the categories *group* and *individual*, the following subdivision can thus be made:

1. Specific characteristic; for example, autism [119], depression [592], and criminals versus students [274] but also baseline blood pressure, hypertensive medication, body mass, smoking [625], and alcohol consumption [227]; see also [633] and [85, Chapter

31; p. 732].

2. Psychological traits; for example, personality [57, 188, 354, 362, 441, 676] or empathy [80, 152, 251, 612]).
3. Demographics; for example, age [314, 435, 553], sex/gender [361, 718], nationality [458], ethnics/race [585, Chapter 28], [56, 314, 401, 603, 627], culture [56, 239, 450, 470], socioeconomic status [239, 470], and level of education [676].
4. Activities; for example, office work [316, 681], driving a car [272, 329, 330, 474] or flying a plane, and running [270].

This subdivision is based on current practice with *ASP*; however, possibly it should be altered.

So far, comparisons between research results on *ASP* are mostly made between results of either individuals or groups selected to resemble the general population (cf. Table 2.4). However, user-tailored approaches should be explored as well. In particular, experiences with specific groups can substantially contribute to the further development of *ASP*, as has been seen in other sciences (e.g., biology, psychology, and medicine).

Having said that, the question remains, how to handle this striking variety between people. We propose three approaches that can possibly tackle these problems:

1. Hybrid classification systems [45]. Most often, such architectures incorporate both a (logic-based) reasoning system and a pattern recognition component. To the authors knowledge, so far, this approach has not been applied for *ASP*. It has, however, been applied successfully for speech-based emotion recognition [591].
2. Multi-agent systems and multi-classifier systems [724]. Two approaches within this field could be of interest: 1) Multi-layered architectures, where each layer determines the possible classes to be processed or the classifiers to be chosen for the next layer and 2) An ensemble of classifiers, trained on the same or distinct biosignals and their features. Their outputs are collected into one compound classification, often determined through a voting scheme. For example, Atassi and Esposito (2008) [20] applied a two-layer classification system for speaker independent classification of six emotions, For more information on this topic, we refer to Lam and Suen [370] and Kuncheva [364].
3. Biosignal signatures. Related to schemes that are used in forensics [559] and with functional neuroimaging (e.g., EEG, fMRI, MEG, NIRSI, SPECT, and PET) [85, Chapter 2; p. 34], *ASP* could benefit from personalized profiles or schemes that tailor to a generic profile of people's unique biosignal signatures [172, 464, 726]. Moreover, this approach could be extended to incorporate context information, as is already done in forensics [559]. Biosignal signatures require advanced multi-modal data mining and knowledge discovery strategies, and are related to the Picard et al.'s baseline ma-

trix [524] (see also Chapter 7).

Each of these approaches enables processing of multi-modal data, which allows researchers to incorporate a range of characteristics (e.g., context, personality, and signals possible to record) [464]. This makes them promising for *ASP* applications, even outside the scope of user identification.

10.4 Conclusion

The signal processing guidelines taken together: physical sensing characteristics, temporal aspects, normalization, and context, all need to be taken into account when processing affective signals. Subsequently, the affective signals need to be classified using pattern recognition techniques. For this phase, the guidelines validation, triangulation, and user identification should be taken into account.

The guidelines presented in this chapter were derived from the author's research conducted on *ASP*. Part of this research can be found in this monograph. Careful processing of all issues mentioned in the guidelines should always be warranted as they provide the input for the core of the closed loop model that forms the core of emotion-aware systems (see Figure 1.1). As such, I hope that this concise set of directives will aid future research in *affective computing*. With the next chapter, I will close this monograph.

11

Discussion

Abstract

This chapter will start with a wrap-up of what has been presented in this monograph. Its main contribution will lie in looking back and forth in time. After an introduction, a historical perspective will be taken, which will illustrate the vast amount of knowledge that is frequently ignored in *ASP* research. Subsequently, I will weigh this monograph's contribution to emotion science's 10 hot topics as has been recently identified [236]. After this, *ASP* will be brought back to practice by introducing *affective computing's I/O*. Next, three applications that fit three disciplines of computer science will be unveiled, namely: Human-Computer Interaction (HCI), Artificial Intelligence (AI), and health informatics. It will be posed that the technique is ready to bring these applications to the market. Subsequently, the pros and cons of two possible future application areas (i.e., robot nannies and a digital human model) will be discussed. I will finish this chapter and, hence, this monograph with a brief conclusion.

In order of appearance, this chapter includes parts of:

Broek, E.L. van den, Zwaag, M.D. van der, Healey, J.A., Janssen, J.H., & Westerink, J.H.D.M. (2010). Prerequisites for Affective Signal Processing (ASP) - Part IV. In J. Kim & P. Karjalainen (Eds.), *Proceedings of the 1st International Workshop on Bio-inspired Human-Machine Interfaces and Healthcare Applications – B-Interface 2010*, p. 59–66. January 21, Valencia. Spain;

Broek, E.L. van den, Sluis, F. van der, & Schouten, Th.E. (2010). User-centered digital preservation of multimedia. *European Research Consortium for Informatics and Mathematics (ERCIM) News, No. 80 (January)*, 45–47;

Broek, E.L. van den (2010). Robot nannies: Future or fiction? *Interaction Studies, 11(2)*, 274–282; *and:*

Broek, E.L. van den (2010). Beyond Biometrics. *Procedia Computer Science, 1(1)*, 2505–2513. *[invited]*.

11.1 Introduction

This monograph was divided into five parts: $I.$ a *prologue*, $II.$ *baseline-free ASP* using statistical moments, $III.$ *bi-modal Affective Signal Processing (ASP)* that explored various possible key factors, IV two studies *towards affective computing*, and $V.$ an *epilogue* of which this discussion is the second part. In addition, Appendix A provides additional background information on the statistical techniques used in this monograph.

The first part, the prologue, started with a general introduction and an introduction on this monograph's key concepts: affect (and emotion), *affective computing*, and *Affective Signal Processing (ASP)*. Next, the closed loop model for *affective computing* and *ASP* was introduced, which served as the working model for this monograph. Moreover, the relevance of *ASP* for three branches of computer science (i.e., Human-Computer Interaction (HCI), Artificial Intelligence (AI), and health informatics) was explained. Last, an outline of this monograph was provided. The second and last chapter of the prologue provided a review of *affective computing* and, more in particular, *ASP*. Biosignals received most attention as this was the target modality of the monograph.

The second part of this monograph consisted of two chapters (Chapters 3 and 4) that presented two distinct sets of analyses on the same data set. The analyses differed in their choice of time windows. This way the impact and usage of this parameter for *ASP* was explored. Dynamic real world stimuli (i.e., fragments from movies) were used to induce emotions, instead of less ecologically valid static stimuli. The EMG of three facial muscles was recorded. This is often done to establish a ground truth measurement. In addition, the participants' EDA's were recorded. This is a robust well-documented biosignal that reveals the level of experienced arousal experienced. Baseline-free *ASP* was achieved through the use of statistical moments. The 3rd and 4th order moments (i.e., skewness and kurtosis) of the biosignals revealed hidden signal characteristics that enabled to discriminate very well between four emotional states with up to 62% explained variance.

The third part of this monograph also consisted of two chapters, Chapters 5 and 6. The studies presented in these chapters only differed with respect to the stimuli used. In the first study, Chapter 5, one of, or perhaps *the* reference set for *affective computing* was used: IAPS images. In the second study, Chapter 6, the same set of movie fragments was used as in Chapters 3 and 4. This enabled a comparison of static versus dynamic stimuli and, as such, assessed their validity. Both studies employed a bi-modal *ASP* approach to assess affective processes, including ECG as biosignal as well as speech. To the best of the author's knowledge, in this context, this combination has so far only been explored by Kim and colleagues [336, 337, 339, 340]. Both studies also explored a range of issues important to *ASP*, namely: emotion models, environment, personality traits, and demographics. Surprisingly, some of these issues were shown to have little influence on *ASP* (e.g., the personality trait extroversion and demographics). In contrast, other issues (e.g., environment and gender) were shown to be of influence. Up to 90% of variance in the data was explained. Moreover, with both studies more support was found for the valence-arousal model than for basic emotions.

The fourth part consisted of three chapters that presented work bringing us further towards *affective computing*. The first chapter, Chapter 7, presented the execution of the complete signal processing + pattern recognition processing pipeline, see also Section 1.5. In the quest for an optimal processing pipeline, several signal processing aspects and classification methods were explored. The second chapter, Chapter 8, assessed the use of the speech signal as affective signal. The study's aim was to explore the feasibility of a speech-based Computer-Aided Diagnosis (CAD) for mental health care. The study consisted of two experiments, one well controlled and one open, in which patients with a post-traumatic stress disorder (PTSD) participated. For both experiments, a model was developed that explained a significant amount of variance. In the third chapter, Chapter 9, the data of Chapter 8 was used to execute the complete signal processing + pattern recognition processing pipeline (cf. Chapter 7). As such, this chapter explores the feasibility of the envisioned *ASP*-based CAD for mental health care. I concluded that both from a clinical and from an engineering point of view, *affective computing* seems to be within reach.

The fifth part, the epilogue, consists of two parts: the discussion you are currently reading and a set of guidelines for *ASP*, which was presented in the previous chapter. This guideline chapter presented the lessons learned while working on the research reported in this monograph. These guidelines indicated possible problems, presented solutions for them, and provided research directives for *affective computing*. As such, this was perhaps the most important chapter of this monograph.

The remainder of this discussion will look back and forth in time. In Section 11.2, I will stress that we should go back to the basics and learn from the field's research history. The reason for this is simple: energy spent on reinvention is wasted. In Sections 11.5 and 11.6, I will go from theory to practice and present some applications that could be realized with the current state-of-art *ASP* as presented in this monograph. Additionally, I will touch upon some of the ethical aspects of these applications. I will end this monograph with a brief conclusion in Section 11.7.

11.2 Historical reflection

Although a lot of knowledge on emotions has been gained over the last centuries [22, Chapter 1], researchers tend to ignore this to a great extent and to stick to some relatively recent theories; for example, the valence and arousal model or the approach avoidance model.

This holds in particular for *ASP* and *affective computing*, where an engineering approach is dominant and a theoretical framework is considered of lesser importance [680]. Consequently, for most engineering approaches, the valence-arousal model is applied as a default option, without considering other possibilities. Nonetheless, a higher awareness of other theories can heighten the understanding and, with that, the success of *ASP*.

It is far beyond the scope of this monograph to provide a complete overview of all of the literature relevant for *ASP* and affective *affective computing*. For such an overview, I refer to the various handbooks and review papers on emotions, affective sciences, and affective neuroscience [16, 52, 72, 139, 144, 208, 209, 238, 396, 492, 535, 573, 582, 631]. In this section, I will touch upon some of the major works on emotion research which originate from medicine, biology, physiology, and psychology.

Let us start with one of the earliest works on biosignals: *De l'Électricité du corps humain* by M. l'Abbé Bertholon (1780) [366], who was the first who described human biosignals. Nearly a century later Darwin (1872) published his book *The expression of emotions in man and animals* [139]. Subsequently, independently of each other, William James and C.G. Lange revealed their theories on emotions, which were remarkably similar [139]. Consequently, their theories have been merged and were baptized the James-Lange theory.

In a nutshell, the James-Lange theory argues that the perception of our own biosignals *is* the emotion. Consequently, no emotions can be experienced without these biosignals. Two decades after the publication of James' theory, this was seriously challenged by Cannon [90, 91] and Bard [29, 30]. They emphasized the role of subcortical structures (e.g., the thalamus, the hypothalamus, and the amygdala) in experiencing emotions. Their rebuttal on the James-Lange theory was expressed in a theory that was founded on five notions:

1. Compared to a normal situation, experienced emotions are similar when biosignals are omitted; e.g., as with the transection of the spinal cord and vagus nerve.
2. Similar biosignals emerge with all emotions. So, these signals cannot cause distinct emotions.
3. The body's internal organs have fewer sensory nerves than other structures. Hence, people are unaware of their possible biosignals.
4. Generally, biosignals have a long latency period, compared to the time emotional responses are expressed.
5. Drugs that trigger the emergence of biosignals do not necessarily trigger emotions in parallel.

I will now address each of Cannon's notions from the perspective of *ASP*. It is important to consider these notions for current *ASP*, as will become apparent.

To the author's knowledge, the first case that illustrated both theories' weaknesses was that of a patient with a lesion, as denoted in Cannon's first notion. This patient reported: *"Sometimes I act angry when I see some injustice. I yell and cuss and raise hell, because if you don't do it sometimes, I learned people will take advantage of you, but it just doesn't have the heat to it that it used to. It's a mental kind of anger."* [285, p. 151]. On the one hand, this case seems to support the James-Lange theory since the lesion disturbed the patient's biosignals and, in parallel, his emotions have diminished or are even absent. On the other hand, the patient does still report emotions, although of a different kind. If biosignals *are* the emotion how can this be explained then? Can this be attributed to higher level cognition, to reasoning only? If not, this can be considered as support for the Cannon-Bard theory. More than anything else this case once more illustrates the complexity of affective processes as well as the need for user identification, in particular research on special cases (see also Section 10.3.3).

The second notion of the Cannon-Bard theory strikes the essence of *ASP*. It would imply that the quest of *affective computing* is doomed to fail. According to Cannon-Bard, *ASP* is of no use since there are no unique sets of biosignals that map to distinct emotions. Luckily, nowadays, this statement is judged as coarse [139]. However, it is generally acknowledged that it is very hard to apply *ASP* successfully [52]. So, to some extent Cannon has been right.

It was confirmed that the number of sensory nerves differs in distinct structures in human bodies (Cannon's notion 3). So, indeed people's physiological structures determine their internal variations to emotional sensitivity. To make *ASP* even more challenging, there are cross-cultural and ethnic differences in people's patterns of biosignals, as was already shown by Sternbach and Tursky [627, 660] and confirmed repeatedly [585, Chapter 28], [556, 557, 603]. Once more this stresses the need for user identification, as is one of the guidelines proposed in this monograph (see Section 10.3.3).

The fourth notion concerns the latency period of biosignals, which Cannon denoted as being 'long'. Indeed a response time is present with biosignals, which one could denote as being long. Moreover, it varies considerably between the several biosignals used with *ASP*; see also Table 1.1 in Chapter 1. The former is a problem, although in most cases a work around is, to some extent, possible. The latter is possibly even more important to take into account, when conducting *ASP*. Regrettably, this is seldom done. This problem has been addressed as temporal construction in Section 10.2.2.

The fifth and last notion of Cannon is one that has not been addressed so far. It goes beyond biosignals since it concerns the neurochemical aspects of emotions. Although this component of human physiology can indeed have a significant influence on experienced emotions, this falls far beyond the scope of this monograph.

It should be noted that the current general opinion among neuroscientists is that the truth lies somewhere in between the theories of James-Lange and Cannon-Bard [139], as was first suggested by Schachter and Singer [575]. However, the various relations between

Cannon's notions and the set of guidelines presented in Chapter 10, illustrate that these notions, although a century old, are still of interest for current *ASP*. Next, I will take an opposite perspective, not a historical but a state-of-the art perspective, founded on Gross' top 10 of hot topics on emotion research [236, p. 215].

11.3 Hot topics: On the value of this monograph

In this section, I will reflect on both the contributions of this monograph to emotions research and the lagoons it has left unexplored. Recently, in the journal *Emotion Review*, James J. Gross summarized his specific top 10 of hot topics on emotion research [236, p. 215]. Gross' hot topics (indicated in *italics*) provide an excellent resource for a structured reflection on this monograph.

1. *Investigating the antecedents of emotions, moods, and other affective processes.*
   Detailed analyses on antecedents of affect have been conducted in Chapter 4. The results illustrated that biosignals are indeed sensitive, reliable, and discriminating with respect to affective processes. Chapters 8 and 9 presented two studies that employed storytelling and reliving to elicit emotions in PTSD patients. The results revealed different patterns in affective responses, which can be attributed both to the method of elicitation and to the antecedents present in both studies.

2. *Developing new tools for analyzing specific emotion-response components, as well as cross-component coherence.*
   A broad range of techniques and tools have been employed throughout this monograph. Appendix A provides a concise overview of the mathematical background on the statistical and machine learning techniques employed. Chapters 3 and 4 introduced statistical moments, in particular skewness and kurtosis, as new features of biosignals that enable the discrimination between emotions. Chapters 5 and 6 explored the extremely rare combination of speech and the biosignal ECG to assess affective states. It proved to be a powerful combination.

3. *Examining bidirectional relations among emotional and cognitive processes ranging from sensation and perception through judgment and decision making to memory.*
   Throughout this monograph this issue has been mentioned several times but has not been an explicit topic of research. Instead, the studies reported in this monograph aimed to isolate affect, as much as possible, and ignored its interaction with cognitive processes, which are very hard to control in ambulatory real-world practice. This monograph treated this issue as a source of noise that the signal processing + pattern recognition pipeline had to cope with.

4. *Describing the functions of emotion-related processes in everyday life.*
   This quest begs for ethnography [150], which is hardly employed in *affective computing*. Chapters 8 and 9 described research conducted on people who suffer from a severe Post Traumatic Stress Disorder (PTSD), which illustrates the impact the disturbance of emotion-related processes on everyday life can have. However, what mechanisms lay at the foundation of mental disorders (e.g., PTSD) remains largely unknown.

5. *Assessing patterns of stability and change in emotion and emotion regulation over the lifespan, from childhood to older age.*
   Such an endeavor requires longitudinal research. This is often conducted in traditional health care. However, in the context of emotion research longitudinal research is rare [416, 417]. This is a typical concern of bringing research *from lab to life* [674], which needs to be considered when bringing *ASP* technology into our everyday lives [674]; see also Chapter 4.

6. *Examining instructed and spontaneous emotion regulation.*
   This topic has been addressed in Chapters 8 and 9. The instructions and tasks the participating patients received assured spontaneous bursts of emotion and illustrated the lack of regulation of them. However, as Gross [236] implies, such studies are indeed (too) rare and should be encouraged.

7. *Analyzing individual differences in emotion-related processes, with an eye to genetic and epigenetic factors.*
   Genetic and epigenetic factors have not been a topic of research in this monograph. Individual differences, however, have been taken into account. Chapters 3, 4, and 7 were devoted to *baseline-free ASP*. Their results suggested the need for individualization. The follow-up studies presented in Chapters 5–6 indeed unveiled individual differences (i.e., environment, the personality trait neuroticism, and gender). Additionally, Section 10.3.3 denoted many more factors of importance.

8. *Exploring cultural differences and similarities in emotion-related processes.*
   Throughout the literature, culture has been shown to be a factor of influence [56, 239, 450, 470], as was also stated in Chapter 10. However, the factor culture has not been a core topic of research in this monograph.

9. *Exploring conceptual and empirical relations between emotion and emotion regulation, on the one hand, and psychological health outcomes on the other.*
   Health and emotion regulation are in constant interaction; consequently, they are impossible to untangle; see also Section 1.6. Therefore, health informatics was identified as one of the disciplines of computer science for which *ASP* is of the utmost importance. Moreover, this was the reason to conduct the two studies reported in Chapters 8 and 9. These studies unveiled how speech can be used for Computer-Aided Diagnosis (CAD) for mental health care.

10. *Assessing the impact of emotion and emotion regulation processes on physical health outcomes.* In Chapter 1 it was noted that emotions influence our cardiovascular system and, consequently, can even shorten or prolong our life. This monograph did, however, not assess the impact of emotions on physical health.

Taken together, throughout this monograph most of Gross' 10 hot topics [236, p. 215] have been addressed, at least to some extent. However, for most topics, it is evident that a significant body of follow-up research is required to unravel the topics in more detail. Nevertheless, in sharp contrast to Solomon's and Russell's concerns but in line with Gross [236], I would like to say *"future's so bright, I gotta wear shades"* [236, p. 212].

Perhaps more than anything else, this assessment of the current monograph illustrated its relevance and provided the necessary additional reflection upon which I conclude with Gross' words: *"This is more than enough work to keep all of us busy who are interested in emotion, so don those sunglasses and let's get to work!"* [236, p. 215]. Next, we will outline how to do this; however, from another perspective, a computer science perspective.

11.4 Impressions / expressions: *Affective Computing's I/O*

In the introduction of this monograph, I already stated that at a first glance, computer science and affective sciences seem to be worlds apart; however, emotions and computers have become entangled and, in time, will inevitably embrace each other. Computer science and practice employs *input/output (I/O)* operations to characterize its processes. This notion can also be fruitfully utilized for *affective computing* and *ASP*, as I will illustrate here (cf. [210]).

Table 11.1 shows a matrix that provides a characterization of machinery using, what could be, standard *I/O*. Machinery without any *I/O* (i.e., –/–) at all is of no use. In constrast, machinery without either input (i.e., *I*) or output (i.e., *O*) are common practice. However, most of us will assume both input and output (i.e., *I/O*), at least to a certain extent, with most of our machinery. For example, take our standard (office) PC with its output (i.e., at least a video (the screen) and audio) and its input (i.e., at least a keyboard and a pointing device). Emerging branches of science and engineering such as AI, AmI, and *affective computing*, however, aim to redecorate this traditional landscape and provide intuitive *I/O* handling. In the case of *affective computing*, what does this imply?

Computer science's notion of *I/O* operations can also be utilized to divide *affective computing* into four categories. In terms of *affective computing*, the output (*O*) denotes the expression of affect (or emotions) and the input (*I*) denotes the perception, impression, or recognition of affect. This division is adapted from the four cases, as they were identified by Rosalind W. Picard's in her thought-paper, which presented her initial thinking on *affective*

		O	
		no	yes
I	no	–/–	I/–
	yes	–/O	I/O

Table 11.1: A description of the four categories of *affective computing* in terms of computer science's input/output (I/O) operations. In terms of *affective computing*, I/O denotes the expression (O) and the perception, impression, or recognition (I) of affect. This division is adapted from the four cases identified by Rosalind W. Picard [520].

computing [520]. Entities without any affective I/O (i.e., –/–), such as traditional machinery, can be very useful in all situations where emotions hinder instead of help. Entities with only affective O could for example be avatars (e.g., characters in games), consumer products (e.g., a sports car), toys for children, and our TV (see also Section 11.5.1). However, such entities would not know what affective state its user is in and, hence, what affect to show as they would lack the affective I for it. So, as its name, emotion-aware systems, already gives away, a requirement for such systems is affective I.

Throughout this monograph, I have focussed on the affective I that is the percept, impression, or recognition of affect. It has been shown to be complex and promising, in parallel; however, Chapter 10 provided a set of prerequisites *ASP* research and development can hold on to. Following these guidelines, successful *ASP* can be employed. Only affective I is possible. In such cases, the affective I alters other processes (e.g., scheduling breaks for pilots) and no affective O is given but another type of output closes the system (cf. Section 1.5 and see Section 11.5.2). In case of affective I/O, the affective O can follow the affective I immediately or with a (fixed or varying) delay. The affective O can also take various forms, as was already denoted in Section 1.6. Moreover, the person who provides the affective I is not necessarily the person who receives the affective O (see Section 11.5.3).

The theoretical framework concerning affective processes is a topic of continuous debate, as was already argued in Section 1.2. Consequently, an accurate interpretation of affective I and, subsequently, an appropriate affective O is hard to establish. In particular in real-world settings, where several sources of noise will disturb the closed-loop (see Section 1.5), this will be a challenging endeavor. So, currently, it is best to apply simple and robust mechanisms to generate affective O (e.g., on reflex agent level [568]) or slightly more advanced. Moreover, it is not specific states of affect that need to be the target but rather the core affect of the user that needs to be smoothly (and unnoticeably) directed to a target core state [316]; see also Section 10.2.2. The next section will provide a few of such applications.

11.5 Applications: Here and now!

In Part IV of this monograph, I presented the research conducted towards *affective computing*. Moreover, in the previous section I have discussed affective I/O to aid a structured discussion towards closing the loop in real-world practice. However, this did not bring us to the development of consumer applications. That is what I will do here and now! In line with affective *I/O*, as outlined in Section 11.4, the two golden rules that secure such endeavors are: control the complexity and follow the guidelines (see Chapter 10).

One of the main rationales behind the applications that will be presented is that the influencing algorithm of the closed loop system (see Figure 1.1) is kept as simple as possible. This suggestion stems from the idea that *ASP* can never be entirely based on psychological changes. As has been discussed in Chapters 2 and 10, many factors outside people's emotional state can contaminate affective signals. A pragmatic solution for this problem can be to express the goals of *ASP* in terms of the affective signals themselves [615], instead of in terms of emotional states. This approach has also been baptized the physiology-driven perspective [654, 680]. With this perspective in mind, I will now present three possible consumer products, one in each discipline of computer science discussed in Chapter 1: HCI, AI, and health informatics.

11.5.1 TV experience

HCI or better human media interaction is part of everyday life. Not only do we interact with our PC but, for example, also with our television [715]. However, as will be illustrated here, human media interaction will soon stretch far beyond that. About a decade ago, Philips developed Ambient Lighting Technology, which is best known as Ambilight [161]. Using an array of LEDs, , mounted at the back side of the television, Ambilight generates in real time video-content matching light effects around the television. These effects not only reduce eye-fatigue [74], but also enlarge the virtual screen resulting in a more immersive viewing experience [597, 599]. The latter can be understood by considering the characteristics of human peripheral vision.

Using real time analysis of both the video and audio signals [255, 705, 742], Ambilight can be augmented and be used to amplify the content's effect on the viewer's emotions. This would result in a loop similar to that presented in Section 1.5. However, note that this would require the viewer to be connected to the TV with a biosensing device, see Figure 1.1. Moreover, the feedback actuator would be the specifications of the Ambilight and/or the specifications of the audio signals. So, the loop would not be closed but rather open. Such an application is well within reach; part of the loop proposed here has already been developed repeatedly over the last decade; that is, the extraction of emotion-related content from audio and video [254, 255, 681, 705, 742].

In more advanced settings, user identification (see Section 10.3.3) could be employed to construct user profiles and tap into the user's personal space [464]. Moreover, physical characteristics (see Section 10.2.1) and the context (see Section 10.2.4) could be taken into account. Ambient light should be considered as just one example of a next generation of applications, which can be extended to various other domains conveniently, such as our cloths and (ambient) lighting.

11.5.2 Knowledge representations

One of AI's core challenges is knowledge representation, which is traditionally approached from an engineering rather than from a user perspective (cf. Chapter 1). Knowledge representation can play five distinct roles: i) a substitute for an object itself; ii) a set of ontological commitments; iii) a fragmentary theory of intelligent reasoning; iv) a medium for pragmatically efficient computation; and v) a medium of human expression [678].

Although knowledge representation has shown its value in a range of domains, I propose to take it one step further, using W3C's Emotion Markup Language (EmotionML). EmotionML is a 'plug-in' language for : $i)$ manual annotation of data; $ii)$ automatic recognition of emotion-related states from user behavior; and $iii)$ generation of emotion-related system behavior. As such, EmotionML enables a fusion between *ASP* and traditional knowledge representations. Amongst a range of other applications, this enables the digital preservation of our experiences augmented with emotions. For example, not only can I record my son's first words or the first time he is playing with a ball, I could also preserve how my son, my wife, and I felt while this happened. Our affective signals (see Chapters 2-7) could be recorded, processed, and filed using EmotionML. In parallel, our affective signals could also be filed as raw signals. In the future, this would perhaps enable advanced digital human modeling.

11.5.3 Computer-Aided Diagnosis (CAD)

As was already stressed in the introduction (Chapter 1) of this monograph, emotions also have their impact on our health [326]. They influence our emotional and mental well-being directly but, as such, also indirectly have their impact on our physical health. Consequently, *ASP* should be considered as an important part of health informatics. In Chapter 8, two models were developed that relate people's stress level to their speech signal. These two models can serve as a springboard for the development of Computer-Aided Diagnosis (CAD), which can serve as a second opinion for a therapist.

As was shown in Chapters 5 and 6, when the application area allows it, other biosignals can conveniently be combined with a speech signal. A combination of biosignals would improve the predictive power of the model, as was already discussed in Chapter 10. The models developed in Chapter 8 were tailored to PTSD patients. Follow-up research could either aim at other specific groups of patients or, preferably, at a generic modeling template that could be conveniently tailored to the patient group at hand. In all cases, the closed loop model can be employed, where the biosensor is the speech signal; hence, an indirect biosignal (see Figure 1.1). The (bio)feedback actuator can consist of (a combination of) a variety of elements among which are therapy session, physical or mental exercises, and the use of medicine. Further, please note that such CAD can also be brought closer to home. A PC could execute the models as presented in Chapter 8 without any problem, even in real time.

11.6 Visions of the future

The previous section discussed applications for which the knowledge and techniques needed are (near to) available. This section stretches way beyond that and looks into two possible future applications of *ASP*. These two applications touch upon various branches of computer science. Moreover, their possible implications will also force us to consider the ethical aspects of *ASP*.

11.6.1 Robot nannies

It was 2010, just one year ago, Sharkey and Sharkey [600] stated: "*. . . robot manufacturers in South Korea and Japan are racing to fulfil that dream with affordable robot "nannies". . . . By extrapolating from ongoing developments in other areas of robotics, we can get a reasonable idea of the facilities that childcare robots could have available to them over the next 5 to 15 years.* " [600, p. 161]. If this estimate is (near to) correct, robot nannies could already take care of my future grandchildren or even my children. This emphasizes that this vision of the future is near, very near, within one generation from now! Will this vision truly be our future or will it remain science fiction (for a while longer)?

Crucial in the development of children is infant attachment; that is "*the deep emotional connection that an infant forms with his or her primary caregiver, often the mother.*" [600, p. 174]. For children's development, a good attachment is crucial. As Sharkey and Sharkey [600, p. 174] also state, "*Responding appropriately to an infant's cues requires a sensitive and subtle understanding of the infant's needs.*" This is also known as maternal sensitivity [166]. The seminal work of Harlow and colleagues [258–260] on baby monkeys already showed the importance of attachment for baby's. Moreover, it enabled the identification of problems that emerge with insecure attachment. To establish maternal sensitivity a tailored emotional connection is required. Par excellence, robot nannies emphasize the need for AI to adopt models of human emotions, as has already been stressed many times throughout this monograph (in particular, see Chapter 1).

The case of robot nannies shows strong resemblances with AI's Intelligent Tutoring Systems (ITS) [12, 621], a branch of expert systems that was vivid 25 years ago. ITS used models of children (i.e., the students aimed to use the ITS) extensively. However, these were tailored towards a specific (sub)domain, whereas robot nannies would need (more) generic models. This is precisely what ITS failed to deliver. Moreover, much more than with ITS, robot nannies should connect emotionally with the children they care for. *ASP* could facilitate such a connection, as has been shown throughout this monograph. However, this adds even more to the complexity of robot nannies, as *ASP* would need to be employed in the noisy real world; in particular, see Chapters 2 and 10 for a discussion on this. Moreover, this gives rise to the ethical aspects of future technology such as robot nannies and AmI [13].

Given the possibly wide implication of *ASP*-augmented technology such as robot nannies and AmI on our daily lives, ethical issues require a significant amount of attention, not after but *before* theory has become practice [200, 280, 525]. Regrettably, it has to be concluded that the resources that are provided for ethical considerations are scarce. Luckily, progress of science and engineering on emerging technology such as robot nannies, AmI, and *ASP* takes considerable time (cf. [89], [195], and [609]). Perhaps, this provides ethics the time to catch up. However, considering humanity's history, I am afraid the chance that ethics will take a position in the forefront of science's landscape is futile.

11.6.2 Digital Human Model

Throughout this monograph it has been noted repeatedly that, if anything, *ASP* needs a holistic approach. However, this has been noted before in other areas of application, amongst which biometrics. In this section, I will link *ASP* and biometrics to each other (cf. [326, 726]) with the aim to work towards a Digital Human Model (DHM). However, before touching upon this aim, I will provide a brief introduction to biometrics.

Four decades ago, IBM envisioned the identification of persons (ID) by machines [149]. IBM stated that this could be achieved through: i) something the user knows or memorizes, ii) something the user carries, and/or iii) a personal physical characteristic. From this concept, the field of biometrics* emerged; that is, "*...the science of establishing the identity of an individual based on the physical, chemical or behavioral attributes of the person.*" [309, p. 1]. Essentially, biometrics is a signal processing + pattern recognition problem [309], just like *ASP*; see Chapter 1. It can be applied to either verify or identify an ID, which can be defined as:

$$I_x = \begin{cases} I_n \text{ if } \max_n\{D(I_x, I_n)\} < T \\ I_x \text{ otherwise} \end{cases} \quad (11.1)$$

where I_x is the representation (e.g., a vector) of an unidentified person, his bioprofile. I_n is

*Biometrics is derived from the Greek language and means: life measuring.

the n^{th} sample from the database (DB). D is a distance metric (e.g., a Minkowsky or quadratic metric) [507] and T is a threshold. Note that in the case Eq. 11.1 results in $I_x = I_x$, the person remains unidentified after the DB is consulted. In case of verification of persons, $1 : 1$ matching is applied. So, the DB, as depicted in Eq. 11.1, contains one profile. Then, $\max_n\{D(I_x, I_n)\} < T$ becomes $D(I_x, I_n) < T$. In practice, frequently a way in between $1 : 1$ and $1 : n$ matching is employed. The bioprofiles that are matched can be built based on:

1. Behavioral attributes; for example, signature, keystroke dynamics, and gait;
2. Physical attributes: fingerprint, iris and retina, facial image and facial thermogram, geometrical features of the face; for example, ear and nose [459] and geometrical features of the hand and feet [175];
3. Other: audio-based (e.g., voice), chemical attributes (e.g., odor), and DNA,

which can be classified using a taxonomy on biometrics [309]. As is illustrated by this enumeration, bioprofiles share various signals with *ASP*; for example ECG [300, 345] and EEG [436]. Hence, not surprisingly, biometric taxonomy shares dimensions with *ASP* (cf. Chaper 10), namely: obtrusiveness, user acceptance, overt versus covert, validity (e.g., the reliability and discriminative power compared with the other signals applied).

ASP can reveal psychological aspects of a person alongside to physiological aspects and as such become a new class of biometrics. From multi-modal (traditional) biometrics, a range of behavioral, chemical, and physical characteristics can be derived. Together *ASP* and biometrics can lay the foundation for a digital human model (DHM; cf. [172, Chapters 16 and 35] and e.g., [726]). On the one hand, a DHM can be seen as the ultimate model for biometrics. On the other hand, a DHM satisfies *ASP*'s need for a holistic approach. The development of a DHM is a vision that will not be easily met.

Building DHM is a process in which law and ethics will claim their place too (cf. [13, 200, 280, 525]). Law considerations comprise: i) rules of privacy, ii) the constitutional background, and iii) privacy under law, including physical, decisional, and information privacy [309, Ch. 18]. People cannot prevent personal information (e.g., biosignals) from being collected by various actors. Therefore, "*several security measures are implemented on servers to minimize the possibility of a privacy violation. Unfortunately, even the most well defended servers are subject to attacks and, however much one trusts a hosting organism/company, such trust does not last forever.*" [11]. One of the possible solutions would be a "*degradation model, where sensitive data undergoes a progressive and irreversible degradation from an accurate state at collection time, to intermediate but still informative fuzzy states, to complete disappearance. We introduce the data degradation model and identify related technical challenges and open issues.*" [11]. More even than with traditional biometrics, DHM requires an adequate handling of this issue.

Ethical considerations emerge from the notion that DHM would enable a much broader information collection than solely a person's (rational) ID. One of the ethical issues is that biometrics introduces the risk of social exclusion [719], which would increase with the introduction of biosignal-based biometrics, as it enables the extraction of much more information than solely traditional biometric data. This makes the balance between intelligence (e.g., AmI) and privacy even more sensitive than with traditional biometrics (cf. [684]). Although there is still a long way to go, it will be interesting to see whether biometrics and *ASP* will indeed merge and evolve to a DHM and if so, what consequences this will have for our lives. Human dignity should be a leading denominator in future research on both DHM and *ASP* [113, 490], perhaps even more than anything else.

11.7 Conclusion

ASP should be considered as a crucial element for HCI, AI, and health informatics. While machines evolve rapidly, incorporating more and more sensors, and receiving more and more autonomy, the interaction with their users has become more delicate than ever before. Users are increasingly starting to demand that computing devices should understand them. Bringing affective processes into AI is said to be the field's missing link. *ASP* can play a crucial role in this process. *ASP*-based technology will prove to be invaluable in supporting our health and well-being. However, the field's progress lays far behind science's (initial) expectations and results are disappointing. This monograph has explored several of *ASP*'s dimensions and as such contributed to the existing body of knowledge. Additionally, a set of guidelines has been presented to provide a concise set of research methods and standards to the field of *ASP*. I hope these guidelines may boost the field's progress.

With this monograph representing the work of just a few years, only a humble step has been made. But now that *ASP* has both academics' and industry's attention, its progress will be accelerated. Perhaps with the lessons learned throughout this monograph and the guidelines it provides, this book will become a reference for *ASP* and aid its research and development. Then, in time, its progress will prove difficult to stop and a new landscape will arise for humanity in which even ethical concerns will need to be redefined to retain their value. Today this may all sound like science fiction; however, there will be a tomorrow in which it will not, in which our being *will* be redefined and perhaps this tomorrow will come even sooner than we may now expect.

Undoubtedly, in time, the progress of *ASP* will prove to be difficult to stop. *ASP* will (possibly) unnoticeably penetrate our everyday lives. Agents and avatars augmented with *ASP* will support us in our work, will increase our level of mindfulness, and improve the quality or our lives. And, as envisioned, ASP will prove to be *the essential key* in the fusion of man and his technology. More importantly, hopefully it can also serve as an interface to help people understanding each other, to help them see what they have in common instead of staring at their differences.

BIBLIOGRAPHY

Bibliography

[1] E. Aarts. Ambient intelligence: Vision of our future. *IEEE Multimedia*, 11(1):12–19, 2004.

[2] J. J. Abbott, Z. Nagy, F. Beyeler, and B. J. Nelson. Robotics in the small, Part I: Microrobotics. *IEEE Robotics & Automation Magazine*, 14(2):92–103, 2007.

[3] D. P. Agrawal. Designing wireless sensor networks: From theory to applications. *Central European Journal of Computer Science*, 1(1):2–18, 2011.

[4] D. W. Aha, D. Kibler, and M. K. Albert. Instance-based learning algorithms. *Machine Learning*, 6(1):37–66, 1991.

[5] H. Alemdar and C. Ersoy. Wireless sensor networks for healthcare: A survey. *Computer Networks*, 54(15):2688–2710, 2010.

[6] D. Alonso-Ríos, A. Vázquez-García, E. Mosqueira-Rey, and V. Moret-Bonillo. A context-of-use taxonomy for usability studies. *International Journal of Human-Computer Interaction*, 26(10):941–970, 2010.

[7] E. Alpaydýn. *Introduction to machine learning*. Adaptive Computation and Machine Learning. Cambridge, MA, USA: The MIT Press, 2nd edition, 2010.

[8] G. W. Alpers, D. Adolph, and P. Pauli. Emotional scenes and facial expressions elicit different psychophysiological responses. *International Journal of Psychophysiology*, 80(3):173–181, 2011.

[9] American Psychiatric Association. *DSM-IV-TR: Diagnostic and Statistical Manual of Mental Disorders*. Washington, DC, USA: American Psychiatric Publishing, Inc., 4 (Text Revision) edition, 2000.

[10] O. Amft and P. Lukowicz. From backpacks to smartphones: Past, present, and future of wearable computers. *IEEE Pervasive Computing*, 8(3):8–13, 2009.

[11] N. Anciaux, L. Bouganim, H. van Heerde, P. Pucheral, and P. M. G. Apers. InstantDB: Enforcing timely degradation of sensitive data. In *Proceedings of the 2008 IEEE 24th International Conference on Data Engineering (ICDE'2008)*, pages 1373–1375, Cancún, México, April 7–12 2008. IEEE Computer Society.

[12] J. R. Anderson, C. F. Boyle, and B. J. Reiser. Intelligent Tutoring Systems. *Science*, 228(4698):456–462, 1985.

[13] M. Anderson and S. L. Anderson. Machine ethics: Creating an ethical intelligent agent. *AI Magazine*, 28(4):15–26, 2007.

[14] B. M. Appelhans and L. J. Luecken. Heart rate variability as an index of regulated emotional responding. *Review of General Psychology*, 10(3):229–240, 2006.

[15] M. A. Arbib and J.-M. Fellous. Emotions: from brain to robot. *TRENDS in Cognitive Sciences*, 8(12):554–561, 2004.

[16] M. B. Arnold. *Emotion and Personality*, volume I (Psychological Aspects) and II (Neurological and Physiological Aspects). New York, NY, USA: Columbia University Press, 1st edition, 1960.

[17] B. Arnrich, O. Mayora, J. Bardram, and G. Tröster. Pervasive healthcare: Paving the way for a pervasive, user-centered and preventive healthcare model. *Methods of Information in Medicine*, 49(1):67–73, 2010.

[18] B. Arnrich, C. Setz, R. La Marca, G. Troster, and U. Ehlert. What does your chair know about your stress level? *IEEE Transactions on Information Technology in Biomedicine*, 14(2):207–214, 2010.

[19] H. H. Asada, P. Shaltis, A. Reisner, S. Rhee, and R. C. Hutchinson. Mobile monitoring with wearable photoplethysmographic biosensors. *IEEE Engineering in Medicine and Biology Magazine*, 22(3):28–40, 2003.

[20] H. Atassi and A. Esposito. A speaker independent approach to the classification of emotional vocal expressions. In *Proceedings of the 20th IEEE International Conference on Tools with Artificial Intelligence*, volume II, pages 147–152, Dayton, OH, USA, November 3–5 2008. Los Alamitos, CA, USA: IEEE Computer Society.

[21] P. K. Atrey, M. A. Hossain, A. El Saddik, and M. S. Kankanhalli. Multimodal fusion for multimedia analysis: A survey. *Multimedia Systems*, 16(6):345–379, 2010.

[22] R. W. Backs and W. Boucsein. *Engineering Psychophysiology: Issues and applications*. Mahwah, NJ, USA: Lawrence Erlbaum Associates, Inc., Publishers, 2000.

[23] R. W. Backs and W. Boucsein. *Psychophysiology in Digital Human Modeling*, pages 16.1–16.14. Boca Raton, FL, USA: CRC Press / Taylor & Francis, 2009.

[24] G. S. Bahr and R. A. Ford. How and why pop-ups don't work: Pop-up prompted eye movements, user affect and decision making. *Computers in Human Behavior*, 27(2):776–783, 2011.

[25] J. N. Bailenson, E. D. Pontikakis, I. B. Mauss, J. J. Gross, M. E. Jabon, C. A. Hutcherson,

C. Nass, and O. John. Real-time classification of evoked emotions using facial feature tracking and physiological responses. *International Journal of Human-Computer Studies*, 66(5):303–317, 2008.

[26] R. Baltissen and W. Boucsein. Autonomic and subjective reactions to predictable and unpredictable aversive stimuli within a habituation paradigm. *Biological Psychology*, 23(1):68, 1986.

[27] R. Banse and K. R. Scherer. Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology*, 70(3):614–636, 1996.

[28] E. I. Barakova and T. Lourens. Expressing and interpreting emotional movements in social games with robots. *Personal and Ubiquitous Computing*, 14(5):457–467, 2010.

[29] P. Bard. On emotional expression after decortication with some remarks on certain theoretical views, Part I. *Psychological Review*, 41(4):309–329, 1934.

[30] P. Bard. On emotional expression after decortication with some remarks on certain theoretical views, Part II. *Psychological Review*, 41(5):424–449, 1934.

[31] L. F. Barrett. Valence as a basic building block of emotional life. *Journal of Research in Personality*, 40:35–55, 2006.

[32] L. F. Barrett and E. A. Kensinger. Context is routinely encoded during emotion perception. *Psychological Science*, 21(4):595–599, 2010.

[33] P. L. Bartlett. The sample complexity of pattern classification with neural networks: the size of the weights is more important than the size of the network. *IEEE Transactions on Information Theory*, 44(2):525–536, 1998.

[34] D. M. Bates and D. G. Watts. *Nonlinear regression analysis and its applications*. Wiley series in Probability and Mathematical Statistics – Applied Probability and Statistics. New York, NY, USA: John Wiley & Sons, Inc., 1988.

[35] J. Bates. The role of emotions in believable agents. *Communications of the ACM*, 37(7):122–125, 1994.

[36] E. B. Baum and D. Haussler. What size net gives valid generalization? *Neural Computation*, 1(1):151–160, 1989.

[37] S. Baurley, P. Brock, E. Geelhoed, and A. Moore. *Communication-Wear*: User feedback as part of a co-design process. *Lecture Notes in Computer Science (Haptic and Audio Interaction Design)*, 4813:56–68, 2007.

[38] R. Beale and C. Creed. Affective interaction: How emotional agents affect users. *International Journal of Human-Computer Studies*, 67(9):755–776, 2009.

[39] A. Bechara. The role of emotion in decision-making: Evidence from neurological patients with orbitofrontal damage. *Brain and Cognition*, 55(1):30–40, 2004.

[40] G. Ben-Shakhar. Standardization within individuals: A simple method to neutralize individual differences in skin conductance. *Psychophysiology*, 22(3):292–299, 1985.

[41] M. Benovoy, J. R. Cooperstock, and J. Deitcher. Biosignals analysis and its application in a performance setting: Towards the development of an emotional-imaging generator. In *Biosignals 2008: Proceedings of the first International Conference on Biomedical Electronics and Devices*, volume 1, pages 253–258, Funchal, Madeira, Portugal, January 28–31 2008. INSTICC.

[42] C. A. Berg and R. Upchurch. A developmental-contextual model of couples coping with chronic illness across the adult life span. *Psychological Bulletin*, 133(6):920–954, 2007.

[43] G. G. Bernston, J. T. Cacioppo, and K. S. Quigley. Respiratory sinus arrhythmia: Autonomic origins, physiological mechanisms, and psychophysiological implications. *Psychophysiology*, 30(2):183–196, 1993.

[44] G. G. Berntson, J. T. Bigger, D. L. Eckberg, P. Grossman, P. G. Kaufmann, M. Malik, H. N. Nagaraja, S. W. Porges, J. P. Saul, P. H. Stone, and M. W. van der Molen. Heart rate variability: Origins, methods, and interpretive caveats. *Psychophysiology*, 34(6):623–648, 1997.

[45] F. Berzal, J.-C. Cubero, D. Sánchez, and J. M. Serrano. ART: A hybrid classification model. *Machine Learning*, 54(1):67–92, 2004.

[46] B. B. Biesecker and L. Erb. Adaptation to living with a genetic condition or risk: A mini-review. *Clinical Genetics*, 74(5):401–407, 2008.

[47] O. Bimber. Brain-Computer Interfaces. *IEEE Computer*, 41(10):[special issue], 2008.

[48] C. M. Bishop. *Pattern Recognition and Machine Learning*. Information Science and Statistics. New York, NY, USA: Springer Science+Business Media, LLC, 2006.

[49] P. H. Blaney. Affect and memory: A review. *Psychological Bulletin*, 99(2):229–246, 1986.

[50] J. Blascovich and J. Bailenson. *Infinite reality: Avatars, eternal life, new worlds, and the dawn of the virtual revolution*. New York, NY, USA: William Morrow / Harper Collins Publishers, 2011.

[51] B. Boashash. *Time frequency signal analysis and processing: A comprehensive reference*. Oxford, UK: Elsevier Ltd., 2003.

[52] K. Boehner, R. DePaula, P. Dourish, and P. Sengers. How emotion is made and measured. *International Journal of Human-Computer Studies*, 65(4):275–291, 2007.

[53] P. Boersma. Accurate short-term analysis of the fundamental frequency and the harmonics-to- noise ratio of a sampled sound. In *Proceedings of the Institute of Phonetic Sciences*, volume 17, pages 97–110. University of Amsterdam, 1993.

[54] P. P. G. Boersma and D. J. M. Weenink. Praat 4.0.4, 2006. URL: http://www.praat.org [Last accessed on August 22, 2011].

[55] F. A. Boiten, N. H. Frijda, and C. J. E. Wientjes. Emotions and respiratory patterns: Review and critical analysis. *International Journal of Psychophysiology*, 17(2):103–128, 1994.

[56] A. Borders and C. T. H. Liang. Rumination partially mediates the associations between perceived ethnic discrimination, emotional distress, and aggression. *Cultural Diversity and Ethnic Minority Psychology*, 17(2):125–133, 2011.

[57] L. Borghans, B. H. H. Golsteyn, J. Heckman, and J. E. Humphries. Identification problems in personality psychology. *Personality and Individual Differences*, 51(3):315–320, 2011.

[58] J. C. Borod. *The neuropsychology of emotion*. Series in Affective Science. New York, NY, USA: Oxford University Press, Inc., 2000.

[59] R. Bose. Sensor networks motes, smart spaces, and beyond. *IEEE Pervasive Computing*, 8(3):84–90, 2009.

[60] B. E. Boser, I. M. Guyan, and V. N. Vapnik. An training algorithm for optimal margin classifiers. In D. Haussler, editor, *Proceedings of the Fifth Annual ACM Workshop on Computational Learning Theory (COLT)*, pages 144–152, Pittsburgh, Pennsylvania, USA, July 27–29 1992. ACM Press: New York, NY, USA.

[61] W. Bosma and E. André. Exploiting emotions to disambiguate dialog acts. In J. Vanderdonckt, N. J. Nunes, and C. Rich, editors, *Proceedings of the 9th international conference on Intelligent User Interfaces*, pages 85–92, Funchal, Madeira, Portugal, 2004. New York, NY, USA: ACM.

[62] W. Boucsein. *Electrodermal activity*. New York, NY, USA: Plenum Press, 1992.

[63] W. Boucsein, F. Schaefer, M. Kefel, P. Busch, and W. Eisfeld. Objective emotional assessment of tactile hair properties and their modulation by different product worlds. *International Journal of Cosmetic Science*, 24(3):135–150, 2002.

[64] G. H. Bower, S. G. Gilligan, and K. P. Monteiro. Selectivity of learning caused by affective states. *Journal of Experimental Psychology: General*, 110(4):451–473, 1981.

[65] K. W. Bowyer, K. Chang, and P. Flynn. A survey of approaches to three-dimensional face recognition. In J. Kittler, M. Petrou, and M. Nixon, editors, *Proceedings of the 17th IEEE International Conference on Pattern Recognition (ICPR 2004)*, volume 1, pages 358–361, Cambridge, United Kingdom, 2004.

[66] M. M. Bradley, M. Codispoti, D. Sabatinelli, and P. J. Lang. Emotion and motivation II: Sex differences in picture processing. *Emotion*, 1(3):300–319, 2001.

[67] M. M. Bradley and P. J. Lang. Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, 25(1):49–59, 1994.

[68] K. A. Brakke, J. M. Mantock, and K. Fukunaga. Systematic feature extraction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-4(3):291–297, 1982.

[69] C. Breazeal. Toward sociable robots. *Robotics and Autonomous Systems*, 42(3–4):167–175, 2003.

[70] J. Broekens, M. Heerink, and H. Rosendal. The effectiveness of assistive social robots in elderly care: A review. *Gerontechnology Journal*, 8(2):94–103, 2009.

[71] J. Broekens, C. M. Jonker, and J.-J. Ch. Meyer. Affective negotiation support systems. *Journal of Ambient Intelligence and Smart Environments*, 2(2):121–144, 2010.

[72] R. Buck. *Human motivation and emotion*. New York, NY, USA: John Wiley & Sons, Inc., 2nd edition, 1988.

[73] M. Bulacu and L. Schomaker. Text-independent writer identification and verification using textural and allographic features. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(4):701–717, 2007.

[74] J. D. Bullough, Y. Akashi, C. R. Fay, and M. G. Figueiro. Impact of surrounding illumination on visual fatigue and eyestrain while viewing television. *Journal of Applied Sciences*, 6(8):1664–1670, 2006.

[75] H. Bunke and A. Sanfeliu. *Syntactic and Structural Pattern Recognition: Theory and Applications*, volume 7 of *Series in Computer Science*. Singapore, Singapore: World Scientific Publishing Co. Pte. Ltd., 1990.

[76] C. J. C. Burges. A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, 2(2):121–167, 1998.

[77] L. K. Bush, U. Hess, and G. Wolford. Transformations for within-subject designs: A Monte Carlo investigation. *Psychological Bulletin*, 113(3):566–579, 1993.

[78] J. Cacioppo and L. Tassinary. Inferring psychological significance from physiological signals. *American Psychologist*, 45(1):16–28, 1990.

[79] J. T. Cacioppo and G. G. Berntson. Relationship between attitudes and evaluative space: A critical review, with emphasis on the separability of positive and negative substrates. *Psychological Bulletin*, 115(3):401–423, 1994.

[80] J. T. Cacioppo and J. Decety. Social neuroscience: Challenges and opportunities in the study of complex behavior. *Annals of the New York Academy of Sciences*, 1224(1):162–173, 2011.

[81] J. T. Cacioppo and D. D. Dorfman. Waveform movement analysis in psychophysiological research. *Psychological Bulletin*, 102(3):421–438, 1987.

[82] J. T. Cacioppo, B. Marshall-Goodell, and D. D. Dorfman. Skeletal muscular patterning: Topographical analysis of the integrated electromyogram. *Psychophysiology*, 20(3):269–283, 1983.

[83] J. T. Cacioppo and R. E. Petty. *Social psychophysiology: A sourcebook*. New York, NY, USA: The Guilford Press, 1983.

[84] J. T. Cacioppo and R. E. Petty. Electromyographic activity over facial muscle regions can differentiate the valence and intensity of affective reactions. *Journal of Personality and Social Psychology*, 50(2):260–268, 1986.

[85] J. T. Cacioppo, L. G. Tassinary, and G. G. Berntson. *Handbook of Psychophysiology*. New York, NY, USA: Cambridge University Press, 3rd edition, 2007.

[86] H. Cai and Y. Lin. Modeling of operators' emotion and task performance in a virtual driving environment. *International Journal of Human-Computer Studies*, 69(9):571–586, 2011.

[87] R. A. Calvo and S. D'Mello. Affect detection: An interdisciplinary review of models, methods, and their applications. *IEEE Transactions on Affective Computing*, 1(1):18–37, 2010.

[88] R. A. Calvo and S. K. D'Mello. *New perspectives on affect and learning technologies*. Explorations in the Learning Sciences, Instructional Systems and Performance Technologies. New York, NY, USA: Springer Science+Business Media, LLC, 2011.

[89] M. Campbell, A. Joseph Hoane, Jr., and F. h. Hsu. Deep Blue. *Artificial Intelligence*, 134(1–2):57–83, 2002.

[90] W. B. Cannon. *Bodily changes in pain, hunger, fear and rage: An account of recent researches into the function of emotional excitement*. New York, NY, USA: D. Appleton and Company, 1915.

[91] W. B. Cannon. The James-Lange theory of emotion: A critical examination and an alternative theory. *American Journal of Psychology*, 39(3–4):106–124, 1927.

[92] P. Carrera and L. Oceja. Drawing mixed emotions: Sequential or simultaneous experiences? *Cognition & Emotion*, 21(2):422–441, 2007.

[93] A. J. Casson, D. C. Yates abd S. J. M. Smith, and J. S. Duncanand E. Rodriguez-Villegas. Wearable electroencephalography. *IEEE Engineering in Medicine and Biology Magazine*, 29(3):44–56, 2010.

[94] I. Ceaparu, J. Lazar, K. Bessiere, J. Robinson, and B. Shneiderman. Determining causes and severity of end-user frustration. *International Journal of Human-Computer Interaction*, 17(3):333–356, 2004.

[95] G. Chanel, J. J. M. Kierkels, M. Soleymani, and T. Pun. Short-term emotion assessment in a recall paradigm. *International Journal of Human-Computer Studies*, 67(8):607–627, 2009.

[96] C.-C. Chang and C.-J. Lin. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2(3):Article No.: 27, 2011.

[97] C.-Y. Chang, J.-S. Tsai, C.-J. Wang, and P.-C. Chung. Emotion recognition with consideration of facial expressions and physiological signals. In *Proceedings of the IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology*, pages 278–283, Nashville, TN, USA, March 30 – April 2 2009. Piscataway, NJ, USA: IEEE Computational Intelligence Society.

[98] W.-Y. Chang, T.-H. Fang, and Y.-C. Lin. Characterization and fabrication of wireless flexible physiological monitor sensor. *Sensors and Actuators A: Physical*, 143(2):196–203, 2008.

[99] P. Y. K. Chau and P. J.-H. Hu. Information technology acceptance by individual professionals: A model comparison approach. *Decision Sciences*, 32(4):699–719, 2001.

[100] S. Chaudhuri, T. D. Pawar, and S. Duttagupta. *Ambulation analysis in wearable ECG*. New York, NY, USA: Springer Science+Business Media, 2009.

[101] Y. Chen, B. Yang, Q. Meng, Y. Zhao, and A. Abraham. Time-series forecasting using a system of ordinary differential equations. *Information Sciences*, 181(1):106–114, 2011.

[102] B. Cheng and G. Liu. Emotion recognition from surface EMG signal using wavelet transform and neural network. In *Proceedings of The 2nd International Conference on Bioinformatics and Biomedical Engineering (ICBBE)*, pages 1363–1366, Shanghai, China, 16–18 May 2008. Piscataway, NJ, USA: IEEE Press.

[103] Y. M. Chi, T.-P. Jung, and G. Cauwenberghs. Dry-contact and noncontact biopotential electrodes: Methodological review. *IEEE Reviews in Biomedical Engineering*, 3(1):106–119, 2010.

[104] A. Choi and W. Woo. Physiological sensing and feature extraction for emotion recognition by exploiting acupuncture spots. *Lecture Notes in Computer Science (Affective Computing and Intelligent Interaction)*, 3784:590–597, 2005.

[105] I. C. Christie and B. H. Friedman. Autonomic specificity of discrete emotion and dimensions of affective space: A multivariate approach. *International Journal of Psychophysiology*, 51(2):143–153, 2004.

[106] M. Clynes. The non-linear biological dynamics of unidirectional sensitivity illustrated by analog computer analysis, pupillary reflex to light and sound, and heart rate behavior. *Annals of the New York Academy of Sciences*, 98(4):806–845, 1962.

[107] M. Clynes. Biocybernetics of the dynamic communication of emotions and qualities. *Science*, 170(3959):764–765, 1970.

[108] M. Clynes. Sentic cycles: The seven passions at your fingertips. *Psychology Today*, 5(1):58–72, 1972.

[109] M. Clynes. *Sentics: The Touch of the Emotions*. New York, NY, USA: Anchor Press/Doubleday, 1977.

[110] M. Clynes. Evaluation of Sentic theory nullified by misunderstood theory and inferior sound: A reply to Nettelbeck, Henderson, and Willson. *Australian Journal of Psychology*, 41(3):327–337, 1989.

[111] P. Cochrane. A measure of machine intelligence. *Proceedings of the IEEE*, 98(9):1543–1545, 2010.

[112] M. Coeckelbergh. Health care, capabilities, and AI assistive technologies. *Ethical Theory and Moral Practice*, 13(2):181–190, 2010.

[113] M. Coeckelbergh. Human development or human enhancement? a methodological reflection on capabilities and the evaluation of information technologies. *Ethics and Information Technology*, 13(2):81–92, 2011.

[114] A. Cohen and J. 't Hart. On the anatomy of intonation. *Lingua*, 19(2):177–192, 1967.

[115] I. Cohen, N. Sebe, A. Garg, L. S. Chen, and T. S. Huang. Facial expression recognition from video sequences: Temporal and static modeling. *Computer Vision and Image Understanding*, 91(1–2):160–187, 2003.

[116] C. Collet, E. Vernet-Maury, G. Delhomme, and A. Dittmar. Autonomic nervous system response patterns specificity to basic emotions. *Journal of the Autonomic Nervous System*, 62(1–2):45–57, 1997.

[117] R. Collier. Physiological correlates of intonation patterns. *Journal of the Acoustical Society of America*, 58(1):249–255, 1975.

[118] R. J. Compton. The interface between emotion and attention: A review of evidence from psychology and neuroscience. *Behavioral and Cognitive Neuroscience Reviews*, 2(2):115–129, 2003.

[119] K. Conn, C. Liu, N. Sarkar, W. Stone, and Z. Warren. Affect-sensitive assistive intervention technologies for children with Autism: An individual-specific approach. In *Proceedings of the 17th IEEE International Symposium on Robot and Human Interactive Communication*, pages 442–447, Munich, Germany, August 1–3 2008. Piscataway, NJ, USA: IEEE, Inc.

[120] D. J. Cook, J. C. Augusto, and V. R. Jakkula. Ambient intelligence: Technologies, applications, and opportunities. *Pervasive and Mobile Computing*, 5(4):277–298, 2009.

[121] D. J. Cook and W. Z. Song. Ambient intelligence and wearable computing: Sensors on the body, in the home, and beyond. *Journal of Ambient Intelligence and Smart Environments*, 1(2):83–86, 2009.

[122] M. T. Cook and A. Agah. A survey of sketch-based 3-D modeling techniques. *Interacting with Computers*, 21(3):201–211, 2009.

[123] C. L. Cooper and L. A. Pervin. *Personality: Critical concepts in psychology*. Critical concepts in psychology. New York, NY, USA: Routledge, 1st edition, 1998.

[124] D. Côrrea and A. Balbinot. Accelerometry for the motion analysis of the lateral plane of the human body during gait. *Health and Technology*, 1(1):35–46, 2011.

[125] C. Cortes and V. Vapnik. Support-Vector Networks. *Machine Learning*, 20(3):273–297, 1995.

[126] P. T. Costa, Jr. and R. R. McCrae. Influence of extraversion and neuroticism on subjective well-being: Happy and unhappy people. *Journal of Personality and Social Psychology*, 38(4):668–678, 1980.

[127] G. W. Cottrell and J. Metcalfe. *EMPATH: Face, emotion, and gender recognition using holons*, volume 3 (Part IX Applications of *NIPS Proceedings Series*, pages 564–571. San Mateo, CA, USA: Morgan Kaufmann Publishers, Inc., 1991.

[128] D. Cousineau. Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method. *Tutorials in Quantitative Methods for Psychology*, 1(1):42–46, 2005.

[129] T. M. Cover and P. E. Hart. Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, IT-13(1):21–27, 1967.

[130] T. M. Cover and J. M. van Campenhout. On the possible orderings in the measure-

ment selection problem. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-7(9):657–661, 1977.

[131] R. Cowie, E. Douglas-Cowie, N. Tsapatsoulis, G. Votsis, S. Kollias, W. Fellenz, and J. G. Taylor. Emotion recognition in human–computer interaction. *IEEE Signal Processing Magazine*, 18(1):32–80, 2001.

[132] D. A. Craig. How do you feel? interoception: The sense of the physiological condition of the body. *Nature Reviews Neuroscience*, 3(8):655–666, 2002.

[133] J. R. Cram and J. V. Basmalian. *Clinical EMG for Surface Recordings*, volume 2. Nevada City, CA, USA: Clinical Resources, 1990.

[134] J. R. Cram, G. S. Kasman, and J. Holtz. *Introduction to Surface Electromyography*. Jones and Bartlett Publishers, Inc. (Ma), 2nd edition, 2008.

[135] A. S. Crandall and D. J. Cook. Coping with multiple residents in a smart environment. *Journal of Ambient Intelligence and Smart Environments*, 1(4):323–334, 2009.

[136] H. D. Critchley. Electrodermal responses: What happens in the brain. *The Neuroscientist*, 8(2):132–142, 2002.

[137] M. M. Cruz-Cunha, A. J. Tavares, and R. J. Simoes. *Handbook of research on developments in e-health and telemedicine: Technological and social perspectives*. Premier Reference Source. Hershey, PA, USA: Information Science Reference / IGI Global, 2010.

[138] T. R. H. Cutmore and D. A. James. Sensors and sensor systems for psychophysiological monitoring: A review of current trends. *Journal of Psychophysiology*, 21(1):51–71, 2007.

[139] T. Dalgleish, B. D. Dunn, and D. Mobbs. Affective neuroscience: Past, present, and future. *Emotion Review*, 1(4):355–368, 2009.

[140] J. K. Darby. *Speech evaluation in Psychiatry*. New York, NY, USA: Grune & Stratton, Inc., 1981.

[141] A. Darwiche. Bayesian networks. *Communications of the ACM*, 53(12):80–90, 2010.

[142] M. Dash and H. Liu. Feature selection for classification. *Intelligent Data Analysis*, 1(3):131–156, 1997.

[143] I. Daubechies. The wavelet transform, time-frequency localization and signal analysis. *IEEE Transactions on Information Theory*, 36(5):961–1005, 1990.

[144] R. J. Davidson, K. R. Scherer, and H. Hill Goldsmith. *Handbook of affective sciences*. New York, NY, USA: Oxford University Press, 2003.

[145] S. Davies. Still building the Memex. *Communications of the ACM*, 54(2):80–88, 2011.

[146] M. de Berg, M. van Kreveld, and M. Overmars. *Computational Geometry: Algorithms and Applications*. Berlin: Springer-Verlag, 3 edition, 2008.

[147] B. de Gelder. Why bodies? Twelve reasons for including bodily expressions in affective neuroscience. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1535):3475–3484, 2009.

[148] B. de Gelder and J. Vroomen. The perception of emotions by ear and by eye. *Cognition & Emotion*, 14(3):289–311, 2000.

[149] K. de Leeuw and J. Bergstra. *The History of Information Security: A Comprehensive Handbook*. Amsterdam, The Netherlands: Elsevier B.V., 2007.

[150] F. de Waal. *The age of empathy: Nature's lessons for a kinder society*. New York, NY, USA: Harmony Books / Crown Publishing Group, 1st edition, 2009.

[151] S. de Waele, G.-J. de Vries, and M. Jäger. Experiences with adaptive statistical models for biosignals in daily life. In *Proceedings of the IEEE 3rd International Conference on Affective Computing and Intelligent Interaction, ACII*, volume 1, pages 710–715, Amsterdam, The Netherlands, 10–12 September 2009. IEEE Press.

[152] J. Decety and W. Ickes. *The social neuroscience of empathy*. Social Neuroscience Series. Cambridge, MA, USA: A Bradford Book / The MIT Press, 2009.

[153] C. DeLancey. *Passionate engines: What emotions reveal about the mind and artificial intelligence*. New York, NY, USA: Oxford University Press, 2002.

[154] N. K. Denzin. *The research act: A theoretical introduction to sociological methods*. Chicago, IL, USA: Aldine Publishing Co., 1970.

[155] S. Derksen and H. J. Keselman. Backward, forward and stepwise automated subset selection algorithms: Frequency of obtaining authentic and noise variables. *British Journal of Mathematical & Statistical Psychology*, 45(2):265–282, 1992.

[156] L. Devillers, L. Lamel, and I. Vasilescu. Emotion detection in task-oriented spoken dialogues. In *Proceedings of the IEEE International Conference on Multimedia and Expo (ICME)*, volume 3, pages III–549–III–552, Baltimore, MD, USA, July 06–09 2003. Piscataway, NJ, USA: IEEE Signal Processing Society.

[157] L. Devroye. The uniform convergence of nearest neighbor regression function estimators and their application in optimization. *IEEE Transactions on Information Theory*, 24(2):142–151, 1978.

[158] A. K. Dey. Understanding and using context. *Personal and Ubiquitous Computing*, 5(1):4–7, 2001.

[159] C. Di Natale, A. Macagnano, R. Paolesse, E. Tarizzo, A. Mantini, and A. D'Amico. Human skin odor analysis by means of an electronic nose. *Sensors and Actuators B: Chemical*, 65(1–3):216–219, 2000.

[160] N. S. Dias, J. P. Carmo, A. F. da Silva, P. M. Mendes, and J. H. Correia. New dry electrodes based on iridium oxide (IrO) for non-invasive biopotential recordings and stimulation. *Sensors and Actuators A: Physical*, 164(1–2):28–34, 2010.

[161] E. M. A. Diederiks, E. R. Meinders, E. van Lier, and R. H. Peters. Method of and system for controlling an ambient light and lighting unit. Patent Application Publication; Appl. No.: 10/519,059 / Patent. No.: US 7,369,903 B2, 2008.

[162] A. Dijkstra. *A computer model for bimodal sublexical processing*, volume 5, chapter 8 (Part 2: Methods), pages 90–101. Lisse, The Netherlands: Swets & Zeitlinger B. V., 1994.

[163] S. Doberenz, W. T. Roth, E. Wollburg, N. I. Maslowski, and S. Kim. Methodological considerations in ambulatory skin conductance monitoring. *International Journal of Psychophysiology*, 80(2):87–95, 2011.

[164] K. Domschke, S. Stevens, B. Pfleiderer, and A. L. Gerlach. Interoceptive sensitivity in anxiety and anxiety disorders: An overview and integration of neurobiological findings. *Clinical Psychology Review*, 30(1):1–11, 2010.

[165] L. X. Dong and B. J. Nelson. Robotics in the small, Part II: Nanorobotics. *IEEE Robotics & Automation Magazine*, 14(3):111–121, 2007.

[166] W. L. Donovan, L. A. Leavitt, and J. D. Balling. Maternal physiological response to infant signals. *Psychophysiology*, 15(1):68–74, 1978.

[167] J. D. M. Drake and J. P. Callaghan. Elimination of electrocardiogram contamination from electromyogram signals: An evaluation of currently used removal techniques. *Journal of Electromyography and Kinesiology*, 16(2):175–187, 2006.

[168] J. Dubnowski, R. Schafer, and L. Rabiner. Real-time digital hardware pitch detector. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 24(1):2–8, 1976.

[169] K. Ducatel, M. Bogdanowicz, F. Scapolo, J. Leijten, and J.-C. Burgelman. Scenarios for ambient intelligence in 2010. Technical report, European Commission Information Society Technologies Advisory Group (ISTAG), February 2001.

[170] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. New York, NY, USA: John Wiley & Sons, Inc., 2nd edition, 2001.

[171] N. M. Dudukovic, S. DuBrow, and A. D. Wagner. Attention during memory retrieval enhances future remembering. *Memory & Cognition*, 37(7):953–961, 2009.

[172] V. G. Duffy. *Handbook of Digital Human Modeling: Research For Applied Ergonomics And Human Factors Engineering*. Human Factors and Ergonomics. Boca Raton, FL, USA: CRC Press / Taylor & Francis, 2009.

[173] A. C. M. Dumaij and E. C. G. Tijssen. On-line health companion contact among chronically ill in the Netherlands. *Health and Technology*, 1(1):5–23, 2011.

[174] F. Dunbar. *Emotions and bodily changes: A survey of literature on psychosomatic interrelationships 1910–1953*. New York, NY, USA: Columbia University Press, fourth edition, 1954.

[175] N. Duta. A survey of biometric technology based on hand shape. *Pattern Recognition*, 42(11):2797–2806, 2009.

[176] T. Eerola and J. K. Vuoskoski. A comparison of the discrete and dimensional models of emotion in music. *Psychology of Music*, 39(1):18–49, 2011.

[177] B. Efron and R. J. Tibshirani. *An introduction to the bootstrap*, volume 57 of *Monographs on Statistics and Applied Probability*. New York, NY, USA: Chapman & Hall, Inc., 1993.

[178] A. Ehlers, J. Bisson, D. M. Clark, M. Creamer, S. Pilling, D. Richards, P. P. Schnurr, S. Turner, and W. Yule. Do all psychological treatments really work the same in posttraumatic stress disorder? *Clinical Psychology Review*, 30(2):269–276, 2010.

[179] P. Ekman. Facial expression and emotion. *American Psychologist*, 48(4):384–392, 1993.

[180] P. Ekman and W. V. Friesen. *Unmasking the face: A guide to recognizing emotions from facial expressions*. A Spectrum Book. Englewood Cliffs, NJ, USA: Prentice-Hall, Inc., 1975.

[181] P. Ekman, R. W. Levenson, and W. V. Friesen. Autonomic nervous system activity distinguishes between emotions. *Science*, 221(4616):1208–1210, 1983.

[182] M. El Ayadi, M. S. Kamel, and F. Karray. Survey on speech emotion recognition: Features, classification schemes, and databases. *Pattern Recognition*, 44(3):572–587, 2011.

[183] N. H. Eller, J. Kristiansen, and øA. M. Hansen. Long-term effects of psychosocial factors of home and work on biomarkers of stress. *International Journal of Psychophysiology*, 79(2):195–202, 2011.

[184] A. Esposito. The perceptual and cognitive role of visual and auditory channels in conveying emotional information. *Cognitive Computation*, 1(3):268–278, 2009.

[185] I. A. Essa and A. P. Pentland. Facial expression recognition using a dynamic model and motion energy. In *IEEE Proceedings of the Fifth International Conference on Com-*

puter Vision, pages 360–367, Cambridge, MA, USA, June 20–23 1995. IEEE Computer Society Press.

[186] I. A. Essa and A. P. Pentland. Coding, analysis, interpretation, and recognition of facial expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):757–763, 1997.

[187] H. J. Eysenck and S. B. G. Eysenck. *Manual of the Eysenck personality scales (EPS Adult)*. Hodder and Stoughton: London, 1991.

[188] S. B. G. Eysenck. An experimental study of psychogalvanic reflex responses of normal, neurotic, and psychotic subjects. *Journal of Psychosomatic Research*, 1(4):258–272, 1956.

[189] P. Fagerberg, A. Ståhl, and K. Höök. eMoto: emotionally engaging interaction. *Personal and Ubiquitous Computing*, 8(1):377–381, 2004.

[190] J. Fahrenberg, F. Foerster, M. Smeja, and W. Muller. Assessment of posture and motion by multichannel piezoresistive accelerometer recordings. *Psychophysiology*, 34(5):607–612, 1997.

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

[192] B. Fasel and J. Luettin. Automatic facial expression analysis: A survey. *Pattern Recognition*, 36(1):259–275, 2003.

[193] J.-M. Fellous and M. A. Arbib. *Who needs emotions? The brain meets the robot*. Series in Affective Science. New York, NY, USA: Oxford University Press, Inc., 2005.

[194] R. Fernandez and R. W. Picard. Signal processing for recognition of human frustration. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, volume 6 (Multimedia Signal Processing), pages 3773–3776, Seattle, WA, USA, May 12–15 1997. Piscataway, NJ, USA: IEEE, Inc.

[195] D. Ferrucci, E. Brown, J. Chu-Carroll, J. Fan, D. Gondek, A. A. Kalyanpur, A. Lally, J. W. Murdock, E. Nyberg, J. Prager, N. Schlaefer, and C. Welty. Building Watson: An overview of the DeepQA project. *AI Magazine*, 31(3):59–79, 2010.

[196] R. B. Fillingim, D. L. Roth, and E. W. Cook III. The effects of aerobic exercise on cardiovascular, facial EMG, and self-report responses to emotional imagery. *Psychosomatic Medicine*, 54(1):109–120, 1992.

[197] R. A. Fisher. The moments of the distribution for normal samples of measures of departure from normality. *Proceedings of the Royal Society of London Series A: Mathematical and Physical Sciences*, 130(812):16–28, 1930.

[198] M. A. Flaten. Effects of auditory frequency-shifts on zero and below-zero SCR habituation. *Integrative Psychological and Behavior Science*, 24(3):98–101, 1989.

[199] C. Flavián-Blanco, R. Gurrea-Sarasa, and C. Orús-Sanclemente. Analyzing the emotional outcomes of the online search behavior with search engines. *Computers in Human Behavior*, 27(1):540–551, 2011.

[200] L. Floridi. *The Cambridge handbook of information and computer ethics*. Cambridge, UK: Cambridge University Press, 2010.

[201] F. Foerster, M. Smeja, and J. Fahrenberg. Detection of posture and motion by accelerometry: A validation study in ambulatory monitoring. *Computers in Human Behavior*, 15(5):571–583, 1999.

[202] J. R. J. Fontaine, K. R. Scherer, E. B. Roesch, and P. Ellsworth. The world of emotion is not two-dimensional. *Psychological Science*, 18(12):1050–1057, 2007.

[203] D. C. Fowles, M. J. Christie, R. Edelberg, W. W. Grings, D. T. Lykken, and P. H. Venables. Publication recommendations for electrodermal measurements. *Psychophysiology*, 18(3):232–239, 1981.

[204] B. L. Frederickson, R. A. Manusco, C. Branigan, and M. M. Tugade. The undoing effect of positive emotions. *Motivation and Emotion*, 24(4):237–257, 2000.

[205] B. L. Fredrickson. The role of positive emotions in positive psychology: The broaden-and-build theory of positive emotions. *American Psychologist*, 56(3):218–226, 2001.

[206] A. J. Fridlund and J. T. Cacioppo. Guidelines for human electromyographic research. *Psychophysiology*, 23(5):567–589, 1986.

[207] M. Friedewald and O. Raabe. Ubiquitous computing: An overview of technology impacts. *Telematics and Informatics*, 28(2):55–65, 2011.

[208] N. H. Frijda. *The emotions*. Studies in Emotion and Social Interaction. Cambridge, UK: Cambridge University Press, 1986.

[209] N. H. Frijda. *The Laws of Emotion*. Mahwah, NJ, USA: Lawrence Erlbaum Associates, Inc., Publishers, 2007.

[210] N. H. Frijda and J. Swagerman. Can computers feel? theory and design of an emotional system. *Cognition & Emotion*, 1(3):235–257, 1987.

[211] C. O. Fritz, P. E. Morris, and J. J. Richler. Effect size estimates: Current use, calculations, and interpretation. *Journal of Experimental Psychology: General*, [in press].

[212] P. Fromherz. Electrical interfacing of nerve cells and semiconductor chips. *2002*, 3(3):276–284, 2002.

[213] P. Fromherz. *Neuroelectronic Interfacing: Semiconductor Chips with Ion Channels, Nerve Cells, and Brain*, chapter 32, pages 781–810. Berlin, Germany: Wiley-VCH Verlag, 2003.

[214] D. R. Frutiger, B. E. Kratochvil, K. Vollmers, and B. J. Nelson. Small, fast, and under control: Wireless resonant magnetic micro-agents. *International Journal of Robotics Research*, 29(5):613–636, 2009.

[215] K. S. Fu. *Syntactic pattern recognition, Applications*, volume 14 of *Communication and Cybernetics*. Berlin/Heidelberg, Germany: Springer-Verlag, 1977.

[216] K. S. Fu. *Syntactic pattern recognition and applications*. Prentice-Hall Advances in Computing Science and Technology Series. Englewood Cliffs, NJ, USA: Prentice Hall, Inc., 1982.

[217] S. Fu, R. Gutierrez-Osuna, A. Esposito, P. K. Kakumanu, and O. N. Garcia. Audio/visual mapping with cross-modal hidden Markov models. *IEEE Transactions on Multimedia*, 7(2):243–252, 2005.

[218] S. Fukuda. *Emotional engineering: Service development*. London, UK: Springer-Verlag London Limited, 2011.

[219] G. Gainotti. *Emotional behavior and its disorders*, volume 5 of *Handbook of Neuropsychology*. Amsterdam, The Netherlands: Elsevier Science B.V., 2001.

[220] H. Gamboa, F. Silva, H. Silva, and R. Falcão. PLUX – Biosignals Acquisition and Processing, 2011. URL: http://www.plux.info [Last accessed on August 22, 2011].

[221] R. Geenen and F. J. R. van de Vijver. A simple test of the law of initial values. *Psychophysiology*, 30(5):525–530, 1993.

[222] G. H. E. Gendolla. On the impact of mood on behavior: An integrative theory and a review. *Review of General Psychology*, 4(4):378–408, 2000.

[223] G. H. E. Gendolla and K. Brinkman. The role of mood states in self-regulation: Effects on action preferences and resource mobilization. *European Psychologist*, 10(3):187–198, 2005.

[224] N. Gershenfeld, R. Krikorian, and D. Cohen. The Internet of things. *Scientific American*, 291(4):76–81, 2004.

[225] Z. Ghahramani. An introduction to Hidden Markov Models and Bayesian networks. *International Journal of Pattern Recognition and Artificial Intelligence*, 15(1):9–42, 2001.

[226] Z. Ghahramani. Unsupervised learning. *Lecture Notes in Artificial Intelligence (Advanced Lectures in Machine Learning)*, 3176:72–112, 2004.

[227] S. Glautier, J. O'Brien, and J. Dixon. Facial electromyographic (EMG) responses to emotionally significant visual images: Differences between light and heavy drinkers. *Drug and Alcohol Dependence*, 64(3):337–345, 2001.

[228] D. Gökçay and G. Yildirim. *Affective Computing and Interaction: Psychological, Cognitive and Neuroscientific Perspectives.* Premier Reference Source. Hershey, PA, USA: Information Science Reference / IGI Global, 2011.

[229] R. C. Gonzalez and M. G. Thomason. *Syntactic Pattern Recognition: An Introduction*, volume 14 of *Applied Mathematics and Computation.* Addison-Wesley, 1978.

[230] F. Gotoh. Influence of affective valence on working memory processes. *International Journal of Psychology*, 43(1):59–71, 2008.

[231] C. Goutte. Note on free lunches and cross-validation. *Neural Computation*, 9(6):1245–1249, 1997.

[232] J. F. Greden. The burden of recurrent depression: Causes, consequences, and future prospects. *The Journal of Clinical Psychiatry*, 62(S22):5–9, 2001.

[233] J. C. Greene and V. J. Caracelli. Advances in mixed-method evaluation: The challenges and benefits of integrating diverse paradigms. *New Directions for Evaluation*, 1997(74):[special issue], 1997.

[234] S. J. Grey and B. L. Smith. A comparison between commercially available electrode gels and purpose-made gel, in the measurement of electrodermal activity. *Psychophysiology*, 21(5):551–557, 1984.

[235] J. J. Gross. *Emotion regulation: Conceptual foundations*, chapter 1, pages 3–24. New York, NY, USA: The Guilford Press, 2007.

[236] J. J. Gross. The futureŠs so bright, I gotta wear shades. *Emotion Review*, 2(3):212–216, 2010.

[237] J. J. Gross and R. W. Levenson. Emotion elicitation using films. *Cognition and Emotion*, 9(1):87–108, 1995.

[238] P. Grossman and E. W. Taylor. Toward understanding respiratory sinus arrhythmia: Relations to cardiac vagal tone, evolution and biobehavioral functions. *Biological Psychology*, 74(2):263–285, 2007.

[239] I. Grossmann, P. C. Ellsworth, and Y.-y. Hong. Culture, attention, and emotion. *Journal of Experimental Psychology: General*, [in press].

[240] H. Gunes and M. Piccardi. Bi-modal emotion recogntion from expressive face and body gestures. *Journal of Network and Computer Applications*, 30(4):1334–1345, 2007.

[241] H. Gunes and M. Piccardi. Automatic temporal segment detection and affect recognition from face and body display. *IEEE Transactions on Systems, Man, and Cybernetics – Part B: Cybernetics*, 39(1):64–84, 2009.

[242] R. Gutierrez-Osuna, P. K. Kakumanu, A. Esposito, O. N. Garcia, A. Bojórquez, J. L. Castillo, and I. Rudomín. Speech-driven facial animation with realistic dynamics. *IEEE Transactions on Multimedia*, 7(1):33–42, 2005.

[243] I. Guyon and A. Elisseeff. An introduction to variable and feature selection. *Journal of Machine Learning Research*, 3(7–8):1157–1182, 2003.

[244] I. M. Guyon, B. E. Boser, and V. N. Vapnik. *Automatic capacity tuning of very large VC–dimension classifiers*, pages 147–155. San Francisco, CA, USA: Morgan Kaufmann, Denver, Colorado, USA, 1993.

[245] A. C. Guyton and J. E. Hall. *Textbook of medical physiology*. Philadelphia, PA, USA: Elsevier Saunders / Elsevier Inc., 11th edition, 2006.

[246] A. Haag, S. Goronzy, P. Schaich, and J. Williams. Emotion recognition using biosensors: First steps towards an automatic system. *Lecture notes in computer science (Affective Dialogue Systems)*, 3068:36–48, 2004.

[247] A. Haarmann, W. Boucsein, and F. Schaefer. Combining electrodermal responses and cardiovascular measures for probing adaptive automation during simulated flight. *Applied Ergonomics*, 40(6):1026–1040, 2009.

[248] T. Habermas and N. Berger. Retelling everyday emotional events: Condensation, distancing, and closure. *Cognition & Emotion*, 25(2):206–219, 2011.

[249] W. M. Haddad. Closed-loop control for intensive care unit sedation. *Best Practice & Research Clinical Anaesthesiology*, 23(1):95–114, 2009.

[250] C. B. Hadley and D. G. MacKay. Does emotion help or hinder immediate memory? arousal versus priority-binding mechanisms. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32(1):79–88, 2006.

[251] J. Håkansson and H. Montgomery. Empathy as an interpersonal phenomenon. *Journal of Social and Personal Relationships*, 20(3):267–284, 2003.

[252] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten. The WEKA data mining software: An update. *ACM SIGKDD Explorations Newsletter*, 11(1):10–18, 2009.

[253] P. Hall, B. U. Park, and R. J. Samworth. Choice of neighbor order in nearest-neighbor classification. *Annals of Statistics*, 36(5):2135–2152, 2008.

[254] A. Hanjalic. Extracting moods from pictures and sounds: Towards truly personalized TV. *IEEE Signal Processing Magazine*, 23(2):90–100, 2006.

[255] A. Hanjalic and L. Q. Xu. Affective video content representation and modeling. *IEEE Transactions on Multimedia*, 7(1):143–154, 2005.

[256] L. K. Hansen and P. Salamon. Neural network ensembles. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(10):993–1001, 1990.

[257] Y. Hao and R. Foster. Wireless body sensor networks for health-monitoring applications. *Physiological Measurement*, 29(11):R27–R56, 2008.

[258] H. F. Harlow. The nature of love. *American Psychologist*, 13(12):673–685, 1958.

[259] H. F. Harlow and M. K. Harlow. Social deprivation in monkeys. *Scientific American*, 207(5):136–146, 1962.

[260] H. F. Harlow and R. R. Zimmermann. Affectional response in the infant monkey: Orphaned baby monkeys develop a strong and persistent attachment to inanimate surrogate mothers. *Science*, 130(3373):421–432, 1959.

[261] R. Harré. Emotions as cognitive-affective-somatic hybrids. *Emotion Review*, 1(4):294–301, 2009.

[262] F. E. Harrell, Jr. *Regression modeling strategies – with applications to linear models, logistic regression, and survival analysis*. Springer Series in Statistics. New York, NY, USA: Springer-Verlag New York, Inc., 1st; 6th printing edition, 2001.

[263] C. M. Harris and M. R. Weiss. Pitch extraction by computer processing of high-resolution Fourier analysis data. *The Journal of the Acoustical Society of America*, 35(3):339–343, 1963.

[264] M. Hassenzahl and N. Tractinsky. User experience – a research agenda. *Behaviour & Information Technology*, 25(2):91–97, 2006.

[265] C. Hatzfeld, T. A. Kern, and R. Werthschützky. Design and evaluation of a measuring system for human force perception parameters. *Sensors and Actuators A: Physical*, 162(2):202–209, 2010.

[266] S. Haykin. *Neural Networks: A comprehensive foundation*. Delhi, India: Prentice Hall / Pearson Education, Inc., 2nd; 9th Indian reprint edition, 2005.

[267] E. S. Healey, A. Kales, L. J. Monroe, E. O. Bixler, K. Chamberlin, and C. R. Soldatos. Onset of insomnia: Role of life-stress events. *Psychosomatic Medicine*, 43(5):439–451, 1981.

[268] J. Healey and R. Picard. SmartCar: detecting driver stress. In A. Sanfeliu and J. J.

Villanueva, editors, *Proceedings of the 17th IEEE International Conference on Pattern Recognition (ICPR 2000)*, volume 4, pages 4218–4121, Barcelona, Spain, 2000.

[269] J. A. Healey. *Wearable and Automotive systems for affect recognition from physiology*. PhD thesis, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology (MIT), 2000.

[270] J. A. Healey. Affect detection in the real world: Recording and processing physiological signals. In *Proceedings of the IEEE 3rd International Conference on Affective Computing and Intelligent Interaction, ACII*, volume 1, pages 729–734, Amsterdam, The Netherlands, 10–12 September 2009. IEEE Press.

[271] J. A. Healey and R. W. Picard. Digital processing of affective signals. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP98)*, volume 6, pages 3749–3752, Seattle, WA, USA, 1998. IEEE.

[272] J. A. Healey and R. W. Picard. Detecting stress during real-world driving tasks using physiological sensors. *IEEE Transactions on Intelligent Transportation Systems*, 6(2):156–166, 2005.

[273] L. Heath. *Triangulation: Methodology*, pages 15901–15906. Elsevier Science Ltd.: Oxford, UK, 1 edition, 2001.

[274] J. H. Hemming. Electrodermal indices in a selected prison sample and students. *Personality and Individual Differences*, 2(1):37–46, 1981.

[275] B. Herbelin, P. Benzaki, F. Riquier, O. Renault, and D. Thalmann. Using physiological measures for emotional assessment: A computer-aided tool for cognitive and behavioural therapy. In P. Sharkey, R. McCrindle, and D. Brown, editors, *Proceedings of the Fifth International Conference on Disability, Virtual Reality and Associated Technologies*, pages 307–314, Oxford, United Kingdom, 2004.

[276] J. A. Herd. Cardiovascular response to stress in man. *Annual Review in Physiology*, 46:177–185, 1984.

[277] H. J. Hermens, B. Freriks, C. Disselhorst-Klug, and G. Rau. Development of recommendations for SEMG sensors and sensor placement procedures. *Journal of Electromyography and Kinesiology*, 10(5):361–374, 2000.

[278] U. Hess, A. Kappas, G. J. McHugo, R. E. Kleck, and J. T. Lanzetta. An analysis of the encoding and decoding of spontaneous and posed smiles: The use of facial electromyography. *Journal of Nonverbal Behavior*, 13(2):121–137, 1989.

[279] D. M. Hilty, S. L. Marks, D. Urness, P. M. Yellowlees, and T. S. Nesbitt. Clinical and educational telepsychiatry applications: A review. *The Canadian Journal of Psychiatry*, 49(1):12–23, 2004.

[280] K. E. Himma and H. T. Tavani. *The handbook of information and computer ethics*. Hoboken, NJ, USA: John Wiley & Sons, Inc., 2008.

[281] M. Hirano, J. Ohala, and W. Vennard. The function of laryngeal muscles in regulating fundamental frequency and intensity of phonation. *Journal of Speech and Hearing Research*, 12(3):616–628, 1969.

[282] E. Hoenkamp. On the notion of "an information need". *Lecture Notes in Computer Science (Theory of Information Retrieval)*, 5766:354–357, 2009.

[283] E. C. M. Hoenkamp. Unitary operators on the document space. *Journal of the American Society for Information Science and Technology*, 54(4):319–325, 2003.

[284] E. C. M. Hoenkamp, O. Stegeman, and L. R. B. Schomaker. Supporting content retrieval from WWW via "basic level categories". In *Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 311–312, Berkeley, CA, USA, August 15–19 1999. New York, NY, USA: ACM Press.

[285] G. W. Hohnmann. Some effects of spinal cord lesions on experienced emotional feelings. *Psychophysiology*, 3(2):143–156, 1966.

[286] K. Hone. Empathic agents to reduce user frustration: The effects of varying agent characteristics. *Interacting with Computers*, 18(4):227–245, 2006.

[287] F. Hönig, A. Batliner, and E. Nöth. Real-time recognition of the affective user state with physiological signals. In *Proceedings of the Doctoral Consortium of Affective Computing and Intelligent Interaction (ACII)*, pages 1–8, Lisbon, Portugal, September 12–14 2007.

[288] K. Höök. Affective loop experiences: Designing for interactional embodiment. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1535):3575–3584, 2009.

[289] S. A. Hosseini, M. A. Khalilzadeh, and S. Changiz. Emotional stress recognition system for affective computing based on bio-signals. *Journal of Biological Systems*, 18(1):101–114, 2010.

[290] E. J. S. Hovenga, M. R. Kidd, S. Garde, and C. H. L. Cossio. *Health Informatics: An Overview*, volume 151 of *Studies in Technology and Informatics*. Amsterdam, The Netherlands: IOS Press BV, 2010.

[291] C.-N. Huang, C.-H. Chen, and H.-Y. Chung. The review of applications and measurements in facial electromyography. *Journal of Medical and Biological Engineering*, 25(1):15–20, 2005.

[292] T. S. Huang, A. Nijholt, M. Pantic, and A. Pentland. *Artifical Intelligence for Human Computing*, volume 4451 of *Lecture Notes in Artificial Intelligence*. Springer Berlin / Heidelberg, 2007.

[293] E. Hudlicka. *Affective Computing: Theory, methods, and applications*. Chapman and Hall/CRC, 2012.

[294] L. G. Hudzinski and G. S. Lawrence. Significance of emg surface electrode placement models and headache findings. *Headache: The Journal of Head and Face Pain*, 28(1):30–35, 1988.

[295] A. Hyvärinen and E. Oja. Independent component analysis: Algorithms and applications. *Neural Networks*, 13:411–430, 2000.

[296] R. M. H. A. Huis in 't Veld, I. A. Widya, R. G. A. Bults, L. Sandsjö, H. J. Hermens, and M. M. R. Vollenbroek-Hutten. A scenario guideline for designing new teletreatments: A multidisciplinary approach. *Journal of Telemedicine and Telecare*, 16(6):302–307, 2010.

[297] D. Isern, D. Sánchez, and A. Moreno. Agents applied in health care: A review. *International Journal of Medical Informatics*, 79(3):145–166, 2010.

[298] T. Isowa, H. Ohira, and S. Murashima. Reactivity of immune, endocrine and cardiovascular parameters to active and passive acute stress. *Biological Psychology*, 65(2):101–120, 2004.

[299] T. Isowa, H. Ohira, and S. Murashima. Immune, endocrine and cardiovascular responses to controllable and uncontrollable acute stress. *Biological Psychology*, 71(2):202–213, 2006.

[300] S. Israel, J. Irvine, A. Cheng, M. Wiederhold, and B. Wiederhold. ECG to identify individuals. *Pattern Recognition*, 38(1):133–142, 2005.

[301] C. E. Izard. More meanings and more questions for the term "emotion". *Emotion Review*, 2(4):383–385, 2010.

[302] C. E. Izard et al. Special section: On defining emotion. *Emotion Review*, 2(4):363–385, 2010.

[303] K. Izdebski. *Emotions in the human voice, Volume I–III*. San Diego, CA, USA: Plural Publishing, Inc., 2008.

[304] Kettunen J and L. Keltikangas-Järvinen. Intraindividual analysis of instantaneous heart rate variability. *Psychophysiology*, 38(4):659–668, 2001.

[305] N. Jackman, W. Schottstaedt, S. C. McPhal, and S. Wolf. Interaction, emotion and physiologic change. *Journal of Health and Human Behavior*, 4(2):83–87, 1963.

[306] A. Jain and D. Zongker. Feature selection: Evaluation, application, and small sample performance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(2):153–158, 1997.

[307] A. K. Jain. Data clustering: 50 years beyond k-means. *Pattern Recognition Letters*, 31(8):651–666, 2010.

[308] A. K. Jain, R. P. W. Duin, and J. Mao. Statistical pattern recognition: A review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(1):4–37, 2000.

[309] A. K. Jain, P. Flynn, and A. A. Ross. *Handbook on Biometrics*. New York, NY, USA: Springer Science+Business Media, LLC, 2008.

[310] A. K. Jain, M. N. Murty, and P. J. Flynn. Data clustering: a review. *ACM Computing Surveys*, 31(3):264–323, 1999.

[311] W. James. What is an emotion? *MIND*, 9(34):188–205, 1884.

[312] W. James. Review: La pathologie des emotions by Ch. Féré. *The Philosophical Review*, 2(3):333–336, 1893.

[313] A. Jameson, T. O. Nelson, R. J. Leonesio, and L. Narens. The feeling of another person's knowing. *Journal of Memory and Language*, 32:320–335, 1993.

[314] C. L. Janes, V. Hesselbrock, and J. A. Stern. Parental psychophysiology, age, and race as related to electrodermal activity of children. *Psychophysiology*, 15(1):24–34, 1978.

[315] J. H. Janssen, J. N. Bailenson, W. A. IJsselsteijn, and J. H. D. M. Westerink. Intimate heartbeats: Opportunities for affective communication technology. *IEEE Transaction on Affective Computing*, 1(2):72–80, 2010.

[316] J. H. Janssen, E. L. van den Broek, and J. H. D. M. Westerink. Tune in to your emotions: A robust personalized affective music player. *User Modeling and User-Adapted Interaction*, 21/2:[in press], 2011/2.

[317] J. R. Jennings, W. K. Berg, J. S. Hutcheson, P. Obrist, S. Porges, and G. Turpin. Publication guidelines for heart rate studies in man. *Psychophysiology*, 18(3):226–231, 1981.

[318] D. N. Joanes and C. A. Gill. Comparing measures of sample skewness and kurtosis. *The Statistician*, 47(Part 1):183–189, 1998.

[319] I. T. Jolliffe. *Principal Component Analysis*. Springer Series in Statistics. New York, NY, USA: Springer-Verlag New York, Inc., 2nd edition, 2002.

[320] C. M. Jones and T. Troen. Biometric valence and arousal recognition. In B. H.

Thomas, editor, *Proceedings of the Australasian Computer-Human Interaction Conference (OzCHI)*, pages 191–194, Adelaide, Australia, 2007.

[321] V. M. Jones, R. Huis in 't Veld, T. Tonis, R. B. Bults, B. van Beijnum, I. Widya, M. Vollenbroek-Hutten, and H. Hermens. Biosignal and context monitoring: Distributed multimedia applications of body area networks in healthcare. In *Proceedings of the 2008 IEEE 10th Workshop on Multimedia Signal Processing (MMSP)*, pages 820–825, Cairns, QLD, Australia, October 8–10 2008. Piscataway, NJ, USA: IEEE Signal Processing Society.

[322] A. Jovic and N. Bogunovic. Electrocardiogram analysis using a combination of statistical, geometric, and nonlinear heart rate variability features. *Artificial Intelligence in Medicine*, 51(3):175–186, 2011.

[323] Deng Ju-Long. Control problems of grey systems. *Systems & Control Letters*, 1(5):288–294, 1982.

[324] K. Jung, K. I. Kim, and A. K. Jain. Text information extraction in images and video: A survey. *Pattern Recognition*, 37(5):977–997, 2004.

[325] A. Kaikkonen, T. Kallio, A. Kekäläinen, A. Kankainen, and M. Cankar. Usability testing of mobile applications: A comparison between laboratory and field testing. *Journal of Usability Studies*, 1(1):4–16, 2005.

[326] A. Kaklauskas, E. K. Zavadskas, V. Pruskus, A. Vlasenko, L. Bartkiene, R. Paliskiene, L. Zemeckyte, V. Gerstein, G. Dzemyda, and G. Tamulevicius. Recommended biometric stress management system. *Expert Systems with Applications Expert Systems with Applications*, 38(11):14011–14025, 2011.

[327] T. W. Kamarck and W. R. Lovallo. Cardiovascular reactivity to psychological challenge: Conceptual and measurement considerations. *Psychosomatic Medicine*, 65(1):9–21, 2003.

[328] A. Kapoor, W. Burleson, and R. W. Picard. Automatic prediction of frustration. *International Journal of Human-Computer Studies*, 65(8):724–736, 2007.

[329] C. D. Katsis, Y. Goletsis, G. Rigas, and D. I. Fotiadis. A wearable system for the affective monitoring of car racing drivers during simulated conditions. *Transportation Research Part C: Emerging Technologies*, 19(3):541–551, 2011.

[330] C. D. Katsis, N. Katertsidis, G. Ganiatsas, and D. I. Fotiadis. Toward emotion recognition in car-racing drivers: A biosignal processing approach. *IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans*, 38(3):502–512, 2008.

[331] B. Kedem. Spectral analysis and discrimination by zero-crossings. *Proceedings of the IEEE*, 74(11):1477–1493, 1986.

[332] T. D. Kelley and L. N. Long. Deep Blue cannot play checkers: The need for generalized intelligence for mobile robots. *Journal of Robotics*, 2010:ID 5237571–8, 2010.

[333] R. C. Kessler. The effects of stressful life events on depression. *Annual Review of Psychology*, 48(1):191–214, 1997.

[334] E. N. Khalil. Communicating affect in news stories: The case of the lead sentence. *Text & Talk – An Interdisciplinary Journal of Language, Discourse Communication Studies*, 26(3):329–349, 2006.

[335] R. Khosla, T. Nishida, and N. Bianchi-Berthouze. *Context-aware emotion-based multi-agent systems*. London, UK: Springer, 2012.

[336] J. Kim. *Bimodal emotion recognition using speech and physiological changes*, chapter 15, pages 265–280. Vienna, Austria: I-Tech Education and Publishing, 2007.

[337] J. Kim and E. André. Emotion recognition using physiological and speech signal in short-term observation. *Lecture Notes in Computer Science (Perception and Interactive Technologies)*, 4021:53–64, 2006.

[338] J. Kim and E. André. Emotion recognition based on physiological changes in music listening. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(12):2067–2083, 2008.

[339] J. Kim, E. André, M. Rehm, T. Vogt, and J. Wagner. Integrating information from speech and physiological signals to achieve emotional sensitivity. In *Proceedings of the 9th European Conference on Speech Communication and Technology*, pages 809–812, Lisboa, Portugal, September 4–8 2005. Lisboa, Portugal: L2F – Spoken Language Systems Laboratory.

[340] J. Kim, E. André, and T. Vogt. Towards user-independent classification of multimodal signals. In *Proceedings of the IEEE 3rd International Conference on Affective Computing and Intelligent Interaction, ACII*, volume 1, pages 722–728, Amsterdam, The Netherlands, 10–12 September 2009. IEEE Press.

[341] K. H. Kim, S. W. Bang, and S. R. Kim. Development of person-independent emotion recognition system based on multiple physiological signals. In *Proceedings of the Second Joint EMBS–BMES Conference*, volume 1: Bioengineering – Integrative Methodologies, New Technologies, pages 50–51, Houston, Texas, USA, 23–26 October 2002.

[342] K. H. Kim, S. W. Bang, and S. R. Kim. Emotion recognition system using short-term monitoring of physiological signals. *Medical & Biological Engineering & Computing*, 42(3):419–427, 2004.

[343] Y. Kim, C. S. Carver, E. L. Deci, and T. Kasser. Adult attachment and psychological

well-being in cancer caregivers: The mediational role of spouses' motives for caregiving. *Health Psychology*, 27(2(Suppl.)):S144–S154, 2008.

[344] J. Kimchi, B. Polivka, and J. S. Stevenson. Triangulation: Operational definitions. *Nursing Research*, 40(6):364–366, 1991.

[345] S. Kiranyaz, T. Ince, J. Pulkkinen, and M. Gabbouj. Personalized long-term ECG classification: A systematic approach. *Expert Systems with Applications*, 38(4):3220–3226, 2011.

[346] A. Kistler, C. Mariauzouls, and K. von Berlepsch. Fingertip temperature as an indicator for sympathetic responses. *International Journal of Psychophysiology*, 29(1):35–41, 1998.

[347] J. Kittler. Mathematical methods of feature selection in pattern recognition. *International Journal of Man-Machine Studies*, 7(5):609–637, 1975.

[348] J. Kittler, M. Hatef, R. P. W. Duin, and J. Matas. On combining classifiers. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(3):226–239, 1998.

[349] R. E. Kleiger, J. T. Bigger, M. S. Bosner, M. K. Chung, J. R. Cook, L. M. Rolnitzky, R. Steinman, and J. L. Fleiss. Stability over time of variables measuring heart rate variability in normal subjects. *The American Journal of Cardiology*, 68(6):626–630, 1991.

[350] P. R. Kleinginna and A. M. Kleinginna. A categorized list of emotion definitions, with a suggestion for a consensual definition. *Motivation and Emotion*, 5(4):345–379, 1981.

[351] A. Kleinsmith, N. Bianchi-Berthouze, and A. Steed. Automatic recognition of non-acted affective postures. *IEEE Transactions on Systems, Man and Cybernetics, Part B: Cybernetics*, [in press].

[352] R. B. Knapp, J. Kim, and E. André. *Physiological signals and their use in augmenting emotion recognition for human-machine interaction*, pages 133–159. Cognitive Technologies. Berlin/Heidelberg, Germany: Springer-Verlag, 2011.

[353] S. Knapp and L. VandeCreek. *Anxiety disorders: A scientific approach for selecting the most effective treatment*. Sarasota, FL, US: Professional Resource Press/Professional Resource Exchange, Inc, 1994.

[354] M. L. Knight and R. J. Borden. Autonomic and affective reactions of high and low socially-anxious individuals awaiting public performance. *Psychophysiology*, 16(3):209–213, 1979.

[355] R. Kohavi and F. Provost. Glossary of terms. *Machine Learning*, 30(2/3):271–274, 1998.

[356] V. Kolodyazhniy, S. D. Kreibig, J. J. Gross, W. T. Roth, and F. H. Wilhelm. An affective computing approach to physiological emotion specificity: Toward subject-independent and stimulus-independent classification of film-induced emotions. *Psychophysiology*, 48(7):908–922, 2011.

[357] E. A. Konijn and J. F. Hoorn. Some like it bad. Testing a model for perceiving and experiencing fictional characters. *Media Psychology*, 7(2):107–144, 2005.

[358] H. Kopetz. *Real-Time Systems: Design principles for distributed and embedded applications*. Real-Time Systems Series. New York, NY, USA: Springer Science+Business Media, LCC, 2nd edition, 2011.

[359] T. R. Kosten, J. W. Mason, E. L. Giller, R. B. Ostroff, and L. Harkness. Sustained urinary norepinephrine and epinephrine elevation in post-traumatic stress disorder. *Psychoneuroendocrinology*, 12(1):13–20, 1987.

[360] S. D. Kreibig, F. H. Wilhelm, W. T. Roth, and J. J. Gross. Cardiovascular, electrodermal, and respiratory response patterns to fear- and sadness-inducing films. *Psychophysiology*, 44(5):787–806, 2007.

[361] A. M. Kring and A. H. Gordon. Sex differences in emotion: Expression, experience, and physiology. *Journal of Personality and Social Psychology*, 74(3):686–703, 1998.

[362] H. W. Krohne, M. Pieper, N. Knoll, and N. Breimer. The cognitive regulation of emotions: The role of success versus failure experience and coping dispositions. *Cognition & Emotion*, 16(2):217–243, 2002.

[363] J. Krumm. *Ubiquitous Computing Fundamentals*. Boca Raton, FL, USA: Chapman & Hall/CRC Press, 2010.

[364] L. I. Kuncheva. A theoretical study on six classifier fusion techniques. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(2):281–286, 2002.

[365] A. de Jongh L. Mevissen. PTSD and its treatment in people with intellectual disabilities: A review of the literature. *Clinical Psychology Review*, 30(3):308–316, 2010.

[366] M. l'Abbé Bertholon. *De l'Électricité du corps humain*. Lyon, France: Tome Premiere, 1780.

[367] J. I. Lacey. *Somatic response patterning and stress: Some revisions of activation theory*, pages 14–42. The Century Psychology Series. New York, NY, USA: Appleton-Century-Crofts, 1967.

[368] J. I. Lacey, D. E. Bateman, and R. VanLehn. Autonomic response specifity: An experimental study. *Psychosomatic Medicine*, 15(1):8–21, 1953. TL1.

[369] D. R. Ladd, K. E. A. Silverman, F. Tolkmitt, G. Bergmann, and K. R. Scherer. Evidence

for the independent function of intonation contour type, voice quality, and $F0$ range in signalling speaker affect. *Journal of the Acoustical Society of America*, 78(2):435–444, 1985.

[370] L. Lam and C. Y. Suen. Optimal combinations of pattern classifiers. *Pattern Recognition Letters*, 16(9):945–954, 1995.

[371] C. Landis and H. N. Dewick. The electrical phenomena of the skin (psychogalvanic reflex). *Psychological Bulletin*, 26(2):64–119, 1929.

[372] P. J. Lang. The emotion probe: Studies of motivation and attention. *American Psychologist*, 50(5):372–385, 1995.

[373] P. J. Lang, M .M. Bradley, and B. N. Cuthbert. Emotion, motivation, and anxiety: Brain mechanisms and psychophysiology. *Biological Psychiatry*, 44(12):1248–1263, 1998.

[374] P. J. Lang, M .M. Bradley, and B. N. Cuthbert. International affective picture system (IAPS): Affective ratings of pictures and instruction manual. Technical Report A-8, University of Florida, Gainesville, FL, USA, 2008.

[375] P. Langley, J. Bowers, and A. Murray. Principal component analysis as a tool for analysing beat-to-beat changes in ECG features: Application to ECG-derived respiration. *IEEE Transactions on Biomedical Engineering*, 57(4):821–829, 2010.

[376] B. G. Lapatki, R. Oostenveld, J. P. van Dijk, I. E. Jonas, M. J. Zwarts, and D. F. Stegeman. Optimal placement of bipolar surface EMG electrodes in the face based on single motor unit analysis. *Psychophysiology*, 47:299–314, 2010.

[377] B. G. Lapatki, D. F. Stegeman, and I. E. Jonas. A surface EMG electrode for the simultaneous observation of multiple facial muscles. *Journal of Neuroscience Methods*, 123(2):117–128, 2003.

[378] B. G. Lapatki, J. P. van Dijk, I. E. Jonas, M. J. Zwarts, and D. F. Stegeman. A thin, flexible multielectrode grid for high-density surface EMG. *Journal of Applied Physiology*, 96(1):327–336, 2004.

[379] J. T. Larsen, A. P. McGraw, and J. T. Cacioppo. Can people feel happy and sad at the same time? *Journal of Personality and Social Psychology*, 81(4):684–696, 2001.

[380] J. T. Larsen, C. J. Norris, and J. T. Cacioppo. Effects of positive and negative affect on electromyographic activity over zygomaticus major and corrugator supercilii. *Psychophysiology*, 40:776–785, 2003.

[381] B. Latré, B. Braem, I. Moerman, C. Blondia, and P. Demeester. A survey on wireless body area networks. *Wireless Networks*, 17(1):1–18, 2011.

[382] P. Legris, J. Ingham, and P. Collerette. Why do people use information technology?

a critical review of the technology acceptance model. *Information & Management*, 40(3):191–204, 2003.

[383] H. Lei, D. M. Sow, J. S. Davis, G. Banavar, and M. R. Ebling. The design and applications of a context service. *Mobile Computing and Communications Review*, 6(4):45–55, 2002.

[384] R. M. Lemov. *World as laboratory: Experiments with mice, mazes, and men*. New York, NY, USA: Hill and Wang, 2005.

[385] D. Lenior, W. Janssen, M. Neerincx, and K. Schreibers. Human-factors engineering for smart transport: Decision support for car drivers and train traffic controllers. *Applied Ergonomics*, 37(4):479–490, 2006.

[386] E. H. Lenneberg. *Biological foundation of language*. New York, NY, USA: John Wiley & Sons, Inc., 1967.

[387] E. Leon, G. Clarke, V. Callaghan, and F. Sepulveda. Real-time detection of emotional changes for inhabited environments. *Computers & Graphics*, 28(5):635–642, 2004.

[388] E. Leon, G. Clarke, V. Callaghan, and F. Sepulveda. A user-independent real-time emotion recognition system for software agents in domestic environments. *Engineering Applications of Artificial Intelligence*, 20(3):337–345, 2007.

[389] E. Leon, G. Clarke, F. Sepulveda, and V. Callaghan. Real-time physiological emotion detection mechanisms: Effects of exercise and affect intensity. In *Proceedings of the 27th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS)*, pages 4719–4722, Shanghai, China, September 1–4 2005. Piscataway, NJ, USA: IEEE, Inc.

[390] R. W. Levenson, L. L. Carstensen, W. V. Friesen, and P. Ekman. Emotion, physiology, and expression in old age. *Psychology & Aging*, 6(1):28–35, 1991.

[391] R. W. Levenson, P. Ekman, K. Heider, and W. V. Friesen. Emotion and the autonomous nervous system activity in the Minangkabau of West Sumatra. *Journal of Personality and Social Psychology*, 62(6):972–988, 1992.

[392] R. W. Levenson and A. M. Ruef. Empathy: A physiological substrate. *Journal of Personality and Social Psychology*, 63(2):234–246, 1992.

[393] C. Levine, D. Halper, A. Peist, and D. A. Gould. Bridging troubled waters: Family caregivers, transitions, and long-term care. *Health Affairs*, 29(1):116–124, 2010.

[394] M. D. Levine. Feature extraction: A survey. *Proceedings of the IEEE*, 57(8):1391–1407, 1969.

[395] G. Lewicki and G. Marino. Approximation by superpositions of a sigmoidal function. *Zeitschrift für Analysis und ihre Anwendungen*, 22(2):463–470, 2003.

[396] M. Lewis, J. M. Haviland-Jones, and L. F. Barrett. *Handbook of Emotions*. New York, NY, USA: The Guilford Press, 3rd edition, 2008.

[397] A. Lichtenstein, A. Oehme, S. Kupschick, and T. Jürgensohn. Comparing two emotion models for deriving affective states from physiological data. *Lecture Notes in Computer Science (Affect and Emotion in Human-Computer Interaction)*, 4868:35–50, 2008.

[398] P. Lieberman and S. B. Michaels. Some aspects of fundamental frequency and envelope amplitude as related to the emotional content of speech. *The Journal of the Acoustical Society of America*, 34(7):922–927, 1962.

[399] J. J.-J. Lien, T. Kanade, J. F. Cohn, and C.-C. Li. Detection, tracking, and classification of action units in facial expression. *Robotics and Autonomous Systems*, 31(3):131–146, 2000.

[400] K. C. Light, P. A. Obrist, S. A. James, and D. S. Strogatz. Cardiovascular responses to stress: II. relationships to aerobic exercise patterns. *Psychophysiology*, 24(1):79–86, 1987.

[401] K. C. Light, P. A. Obrist, A. Sherwood, S. A. James, and D. S. Strogatz. Effects of race and marginally elevated blood pressure on responses to stress. *Hypertension*, 10(6):555–563, 1987.

[402] R. Likert. A technique for the measurement of attitudes. *Archives of Psychology*, 22(140):5–53, 1932.

[403] C. K. Lim, Z. Luo, I-M. Chen, and S. H. Yeo. Wearable wireless sensing system for capturing human arm motion. *Sensors and Actuators A: Physical*, 166(1):125–132, 2011.

[404] Y. S. Lincoln and E. G. Guba. *Naturalistic inquiry*. Newbury Park, CA, USA: Sage Publications, Inc., 1985.

[405] C. L. Lisetti and F. Nasoz. MAUI: a multimodal affective user interface. In *Proceedings of the tenth ACM international conference on Multimedia*, pages 161–170, Juan les Pins, France, December 01–06 2002. New York, NY, USA: ACM Press.

[406] 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.

[407] C. L. Lisetti and F. Nasoz. Affective intelligent car interfaces with emotion recogni-

tion. In *Proceedings of 11th International Conference on Human-Computer Interaction*, pages [on cd–rom], Las Vegas, Nevada, USA, 2005. Mahwah, NJ, USA: Lawrence Erlbaum Associates, Inc.

[408] G. Littlewort, M. S. Bartlett, I. Fasel, J. Susskind, and J. Movellan. Dynamics of facial expression extracted automatically from video. *Image and Vision Computing*, 24(6):615–625, 2006.

[409] C. Liu, K. Conn, N. Sarkar, and W. Stone. Affect recognition in robot assisted rehabilitation of children with autism spectrum disorder. In *Proceedings of the IEEE International Conference on Robotics and Automation*, pages 1755–1760, Roma, Italy, April 10–14 2007. Piscataway, NJ, USA: IEEE, Inc.

[410] C. Liu, K. Conn, N. Sarkar, and W. Stone. Online affect detection and robot behavior adaptation for intervention of children with Autism. *IEEE Transactions on Robotics*, 24(4):883–896, 2008.

[411] C. Liu, P. Rani, and N. Sarkar. An empirical study of machine learning techniques for affect recognition in human-robot interaction. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 2662–2667, Edmonton, Alberta, Canada, August 2–6 2005. Piscataway, NJ, USA: IEEE Press.

[412] C. Liu, P. Rani, and N. Sarkar. Human-robot interaction using affective cues. In *Proceedings of the 15th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN06)*, pages 285–290, Hatfield, UK, 2006. IEEE Computer Society.

[413] G.-H. Liu and J.-Y. Yang. Image retrieval based on the texton co-occurrence matrix. *Pattern Recognition*, 41(12):3521–3527, 2008.

[414] X. Liu, Y. Zheng, M. Phyu, B. Zhao, M. Je, and X. Yuan. Multiple functional ECG signal is processing for wearable applications of long-term cardiac monitoring. *IEEE Transactions on Biomedical Engineering*, 58(2):380–389, 2011.

[415] S. E. Lively, D. B. Pisoni, W. van Summers, and R. H. Bernacki. Effects of cognitive workload on speech production: Acoustic analyses and perceptual consequences. *Journal of the Acoustical Society of America*, 93(5):2962–2973, 1993.

[416] M. M. Llabre, P. G. Saab, B. E. Hurwitz, N. Schneiderman, C. A. Frame, S. Spitzer, and D. Phillips. The stability of cardiovascular parameters under different behavioral challenges: One-year follow-up. *International Journal of Psychophysiology*, 14(3):241–248, 1983.

[417] M. M. Llabre, P. G. Saab, B. E. Hurwitz, N. Schneiderman, C. A. Frame, S. Spitzer, and D. Phillips. The stability of cardiovascular parameters under different behavioral chal-

lenges: One-year follow-up. *International Journal of Psychophysiology*, 14(3):241–248, 1993.

[418] M. M. Llabre, S. B. Spitzer, P. G. Saab, G. H. Ironson, and N. Schneiderman. The reliability and specificity of delta versus residualized change as a measure of cardiovascular reactivity to behavioral challenges. *Psychophysiology*, 28(6):701–711, 1991.

[419] N. K. Logothetis. What we can do and what we cannot do with fMRI. *Nature*, 453(7197):869–878, 2008.

[420] F. Lotte, M. Congedo, A. Lécuyer, F. Lamarche, and B. Arnaldi. A review of classification algorithms for EEG-based brain-computer interfaces. *Journal of Neural Engineering*, 4(2):R1–R13, 2007.

[421] T. Lourens, R. van Berkel, and E. Barakova. Communicating emotions and mental states to robots in a real time parallel framework using Laban movement analysis. *Robotics and Autonomous Systems*, 58(12):1256–1265, 2010.

[422] R. E. Lucas and B. M. Baird. Extraversion and emotional reactivity. *Journal of Personality and Social Psychology*, 86(3):473–485, 2004.

[423] R. E. Lucas, E. Diener, and R. J. Larsen. *Measuring positive emotions*, volume 39 of *Social Indicators Research Series*, pages 139–155. Dordrecht, The Netherlands: Springer Science+Business Media B.V., 2009.

[424] I. Luengo, E. Navas, and I. Hernáez. Feature analysis and evaluation for automatic emotion identification in speech. *IEEE Transactions on MultiMedia*, 12(6):490–501, 2010.

[425] P. Lukowicz. Wearable computing and artificial intelligence for healthcare applications. *Artificial Intelligence in Medicine*, 42(2):95–98, 2008.

[426] M. Lungarella, F. Lida, J. C. Bongard, and R. Pfeifer. *50 Years of Artificial Intelligence: Essays dedicated to the 50th anniversary of Artificial Intelligence*, volume 4850 of *Lecture Notes in Artificial Intelligence*. Berlin/Heidelberg, Germany: Springer-Verlag, Monte Verità, Ascona, Switzerland, 2007.

[427] D. T. Lykken. Research with twins: The concept of emergenesis. *Psychophysiology*, 19(4):361–372, 1982.

[428] D. T. Lykken. The mechanism of emergenesis. *Genes, Brain and Behavior*, 5(4):306–310, 2006.

[429] D. T. Lykken, W. G. Iacono, K. Haroian, M. Mc Gue, and T. J. Bouchard Jr. Habituation of the skin conductance response to strong stimuli: A twin study. *Psychophysiology*, 25(1):4–15, 1988.

[430] D. T. Lykken, R. Rose, B. Luther, and M. Maley. Correcting for psychophysiological measures for individual differences in range. *Psychological Bulletin*, 66(6):481–484, 1966.

[431] Z. Lynch. The future of emotions and robots. *The Lancet Neurology*, 4(12):803, 2005.

[432] R. G. Lyons. *Understanding Digital Signal Processing*. Upper Saddle River, NJ, USA: Prentice Hall PTR, 2nd edition, 2004.

[433] P. W. Macfarlane, A. Van Oosterom, M. Janse, and P. Kligfield. *Comprehensive Electrocardiology: Theory and Practice in Health and Disease*. London, UK: Springer, 2nd edition, 2009.

[434] M. Malik, J. T. Bigger, A. J. Camm, R. E. Kleiger, A. Malliani, A. J. Moss, and P. J. Schwartz. Heart rate variability: Standards of measurement, physiological interpretation, and clinical use. *European Heart Journal*, 17(3):354–381, 1996.

[435] M. Malik and A. J. Camm. *Heart Rate Variability*. Armonk, NY, USA: Futura Publishing Company, Inc., 1995.

[436] S. Marcel and J. del R. Millán. Person authentication using brainwaves (EEG) and maximum a posteriori model adaptation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(4):743–752, 2007.

[437] I. Martin and J. Rust. Habituation and the structure of the electrodermal system. *Psychophysiology*, 13(6):554–562, 1976.

[438] J. F. Martínez, M. S. Familiar, I. Corredor, A. B. García, S. Bravo, and L. López. Composition and deployment of e-Health services over Wireless Sensor Networks. *Mathematical and Computer Modelling*, 53(3–4):485–503, 2011.

[439] A. Marty. *Untersuchungen zur allgemeinen grundlegung der grammatik und sprachphilosophie*. Halle/Saale, Germany: Niemeyer, 1908.

[440] M. Marwitz and G. Stemmler. On the status of individual response specificity. *Psychophysiology*, 35(1):1–15, 1998.

[441] G. Matthews. Personality and multidimensional arousal: A study of two dimensions of extraversion. *Personality and Individual Differences*, 8(1):9–16, 1987.

[442] K. Matzler, R. Faullant, B. Renzl, and V. Leiter. The relationship between personality traits (extraversion and neuroticism), emotions and customer self-satisfaction. *Innovative Marketing*, 1(2):32–39, 2005.

[443] I. B. Mauss and M. D. Robinson. Measures of emotion: A review. *Cognition and Emotion*, 23(2):209–237, 2009.

[444] S. W. McQuiggan and J. C. Lester. Modeling and evaluating empathy in embodied companion agents. *International Journal of Human-Computer Studies*, 65(4):348–360, 2007.

[445] W. S. Meisel. *Computer-oriented approaches to pattern recognition*, volume 83 of *Mathematics in Science and Engineering*. New York, NY, USA: Academic Press, Inc., 1972.

[446] R. Merletti and P. A. Parker. *Electromyography: Physiology, Engineering, and Non-Invasive Applications*. IEEE Press Series on Biomedical Engineering. Hoboken, NJ, USA: John Wiley & Sons, Inc., 2004.

[447] M. Merri, D. C. Farden, J. G. Mottley, and E. L. Titlebaum. Sampling frequency of the electrocardiogram for spectral analysis of the heart rate variability. *IEEE Transactions on Biomedical Engineering*, 37(1):99–106, 1990.

[448] Merriam-Webster, Incorporated. Merriam-Webster Online. URL: http://www.m-w.com/ [Last accessed on August 22, 2011], 2011.

[449] L. Mesin, R. Merletti, and A. Rainoldi. Surface EMG: The issue of electrode location. *Journal of Electromyography and Kinesiology*, 19(5):719–726, 2009.

[450] B. Mesquita and N. H. Frijda. Cultural variations in emotions: A review. *Psychological Bulletin*, 112(2):179–204, 1992.

[451] P. Mielle, A. Tarrega, E. Sémon, J. Maratray, P. Gorria, J. J. Liodenot, J. Liaboeuf, J.-L. Andrejewski, and C. Salles. From human to artificial mouth, from basics to results. *Sensors and Actuators B: Chemical*, 146(2):440–445, 2010.

[452] J. A. Mikels, B. L. Fredrickson, G. R. Larkin, C. M. Lindberg, S. J. Maglio, and P. A. Reuter-Lorenz. Emotional category data on images from the international affective picture system. *Behavior Research Methods*, 37(4):626–630, 2005.

[453] T. Millon and M. J. Lerner. *Personality and social psychology*, volume 5 of *Handbook of Psychology*. Hoboken, NJ, USA: John Wiley & Sons, Inc., 2003.

[454] M. Minsky. *The Society of Mind*. New York, NY, USA: Simon & Schuster Paperbacks, 1985.

[455] M. Minsky. *The Emotion Machine: Commonsense Thinking, Artificial Intelligence, and the Future of the Human Mind*. New York, NY, USA: Simon & Schuster, 2006.

[456] D. A. Mitchell and P. H. Venables. The relationship of EDA to electrode size. *Psychophysiology*, 17(4):408–412, 1980.

[457] T. M. Mitchell. *Machine Learning*. McGraw-Hill Series in Computer Science. Boston, MA, USA: The McGraw-Hill Companies, Inc., 1997.

[458] Y. Miyamoto, Y. Uchida, and P. C. Ellsworth. Culture and mixed emotions: Co-occurrence of positive and negative emotions in Japan and the United States. *Emotion*, 10(3):404–415, 2001.

[459] A. Moorhouse, A. N. Evans, G. A. Atkinson, J. Sun, and M. L. Smith. The nose on your face may not be so plain: Using the nose as a biometric. In *Proceedings of the 3rd International Conference on Crime Detection and Prevention*, page P03, London, UK, December 03 2009. Stevenage, Herts, UK: The Institution of Engineering and Technology (IET).

[460] D. P. Morales, A. García, E. Castillo, M. A. Carvajal, J. Banqueri, and A. J. Palma. Flexible ECG acquisition system based on analog and digital reconfigurable devices. *Sensors and Actuators A: Physical*, 165(2):261–270, 2011.

[461] S. Mori, C. Y. Suen, and K. Yamamoto. Historical review of OCR research and development. *Proceedings of the IEEE*, 80(7):1029–1058, 1992.

[462] W. N. Morris. *Mood: The frame of mind*. Springer Series in Social Psychology. New York, NY, USA: Springer-Verlag, 1989.

[463] D. Morrison, R. Wang, and L. C. De Silva. Ensemble methods for spoken emotion recognition in call-centres. *Speech Communication*, 49(2):98–112, 2007.

[464] E. Mower, M. J. Matarić, and S. Narayanan. A framework for automatic human emotion classification using emotion profiles. *IEEE Transactions on Audio, Speech, and Language Processing*, 19(5):1057–1070, 2011.

[465] H. Mukerjee. Nearest neighbor regression with heavy-tailed errors. *The Annals of Statistics*, 21(2):681–693, 1993.

[466] L. J. M. Mulder. Measurement and analysis methods of heart rate and respiration for use in applied environments. *Biological Psychology*, 34(2–3):205–236, 1992.

[467] L. J. M. Mulder, C. Dijksterhuis, A. Stuiver, and D. de Waard. Cardiovascular state changes during performance of a simulated ambulance dispatchers' task: Potential use for adaptive support. *Applied Ergonomics*, 40(6):965–977, 2009.

[468] K.-R. Müller, S. Mika, G. Rätsch, K. Tsuda, and B. Schölkopf. An introduction to kernel-based learning algorithms. *IEEE Transactions on Neural Networks*, 12(2):181–201, 2001.

[469] I. R. Murray and J. L. Arnott. Toward the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion. *Journal of the Acoustical Society of America*, 93(2):1097–1108, 1993.

[470] J. Na, I. Grossmann, M. E. W. Varnum, R. Gonzalez, S. Kitayama, and R. E. Nisbett.

When cultural differences are not reducible to individual differences. *Proceedings of the National Academy of Sciences of the United States of America*, 107(14):6192–6197, 2010.

[471] H. Nakashima, H. Aghajan, and J. C. Augusto. *Handbook of Ambient Intelligence and Smart Environments*. New York, NY, USA: Springer Science+Business Media, LLC, 2010.

[472] B. A. Nardi, D. J. Schiano, M. Gumbrecht, and L. Swartz. Why we blog. *Communications of the ACM*, 47(12):41–46, 2004.

[473] F. Nasoz, K. Alvarez, C. L. Lisetti, and N. Finkelstein. Emotion recognition from physiological signals for presence technologies. *International Journal of Cognition, Technology and Work*, 6(1):4–14, 2003.

[474] F. Nasoz, C. L. Lisetti, and A. V. Vasilakos. Affectively intelligent and adaptive car interfaces. *Information Sciences*, 180(20):3817–3836, 2010.

[475] C. I. Nass, Y. Moon, B. J. Fogg, B. Reeves, and D. C. F. Dryer. Can computer personalities be human personalities? *International Journal of Human-Computer Studies*, 43(2):223–239, 1995.

[476] M. D. Naylor, L. H. Aiken, E. T. Kurtzman, D. M. Olds, and K. B. Hirschman. The importance of transitional care in achieving health reform. *Health Affairs*, 30(4):746–754, 2011.

[477] D. A. Neamen. *Microelectronics: Circuit analysis and design*. New York, NY, USA: McGraw-Hill Higher Education, 4th edition, 2010.

[478] U. Neisser. The imitation of man by machine – The view that machines will think as man does reveals misunderstanding of the nature of human thought. *Science*, 139(3551):193–197, 1963.

[479] B. J. Nelson, I. K. Kaliakatsos, and J. J. Abbott. Microrobots for minimally invasive medicine. *Annual Review of Biomedical Engineering*, 12:55–85, 2010.

[480] T. Nettelbeck, C. Henderson, and R. J. Willson. Communicating emotion through sound: An evaluation of Clynes' theory of Sentics. *Australian Journal of Psychology*, 41(1):25–36, 1989.

[481] S. A. Neumann and S. R. Waldsein. Similar patterns of cardiovascular response during emotional activation as a function of affective valence and arousal and gender. *Journal of Psychosomatic Research*, 50(5):245–253, 2001.

[482] J. A. Nevin and G. S. Reynolds. *The study of behavior: Learning, motivation, emotion, and instinct*. Glenview, IL, USA: Scott, Foresman and Company, 1973.

[483] M. G. Newman, L. E. Szkodny, S. J. Llera, and A. Przeworski. A review of technology-assisted self-help and minimal contact therapies for anxiety and depression: Is human contact necessary for therapeutic efficacy? *Clinical Psychology Review*, 31(1):89–103, 2011.

[484] P. M. Niedenthal, M. Augustinova, and M. Rychlowska. Body and mind: Zajonc's (re)introduction of the motor system to emotion and cognition. *Emotion Review*, 2(4):340–347, 2010.

[485] A. Nijholt. Google home: Experience, support and re-experience of social home activities. *Information Sciences*, 178(3):612–630, 2008.

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

[487] T. Nishida, L. C. Jain, and C. Faucher. *Modelling machine emotions for realizing intelligence: Foundations and applications*, volume 1 of *Smart innovation, systems and technologies*. Berlin/Heidelberg, Germany: Springer-Verlag, 2010.

[488] R. M. Ba nos, V. Guillen, S. Quero, A. García-Palacios, M. Alcaniz, and C. Botella. A virtual reality system for the treatment of stress-related disorders: A preliminary analysis of efficacy compared to a standard cognitive behavioral program. *International Journal of Human-Computer Studies*, 69(9):602–613, 2011.

[489] D. Novak, J. Ziherl, A. Olenšek, M. Milavec, J. Podobnik, M. Mihelj, and M. Munih. Psychophysiological responses to robotic rehabilitation tasks in stroke. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 18(4):351–361, 2010.

[490] M. C. Nussbaum and A. Sen. *Frontiers of justice: Disability, nationality, species membership*. The Tanner lectures on human values. Cambridge, MA, USA: The Belknap Press / Harvard University Press, 2006.

[491] T. L. Nwe, S. W. Foo, and L. C. De Silva. Speech emotion recognition using hidden Markov models. *Speech Communication*, 41(4):603–623, 2003.

[492] I. Nyklíček, A. Vingerhoets, and M. Zeelenberg. *Emotion regulation and well-being*. New York, NY, USA: Springer Science+Business Media, LLC, 2011.

[493] P. A. Obrist. *Cardiovascular psychophysiology: A perspective*. New York, NY, USA: Plenum Press, 1981.

[494] P. A. Obrist, A. H. Black, J. Brener, and L. V. DiCara. *Cardiovascular Psychophysiology: Current issues in response mechanisms, biofeedback, and methodology*. Chicago, IL, USA: Aldine Publishing Company, 1974.

[495] P. A. Obrist, C. J. Gaebelein, E. S. Teller, A. W. Langer, A. Grignolo, K. C. Light, and J. A. McCubbin. The relationship among heart rate, carotid dP/dt, and blood pressure in humans as a function of the type of stress. *Psychophysiology*, 15(2):102–115, 1978.

[496] P. A. Obrist, K. C. Light, S. A. James, and D. S. Strogatz. Cardiovascular responses to stress: I. measures of myocardial response and relationship to high resting systolic pressure and parental hypertension. *Psychophysiology*, 24(1):65–78, 1987.

[497] T. Ohhashi, M. Sakaguchi, and T. Tsuda. Human perspiration measurement. *Physiological Measurement*, 19(4):449–461, 1998.

[498] R. H. Okada. A critical review of vector electrocardiography. *IRE Transactions on Bio-Medical Electronics*, 10(3):95–98, 1963.

[499] S. Olejnik, J. Li, S. Supattathum, and C. J. Huberty. Multiple testing and statistical power with modified Bonferroni procedures. *Journal of Educational & Behavioral Statistics*, 22(4):389–406, 1997.

[500] J. Or. *Affective Computing*. Vienna, Austria: In-Tech Education and Publishing, 2008.

[501] R. F. Orlikoff and R. J. Baken. Fundamental frequency modulation of the human voice by the heartbeat: Preliminary results and possible mechanisms. *The Journal of the Acoustical Society of America*, 85(2):888–893, 1989.

[502] A. Ortony, G. L. Clore, and A. Collins. *The cognitive structure of emotions*. New York, NY, USA: Cambridge University Press, 1988.

[503] P.-Y. Oudeyer. The production and recognition of emotions in speech: Features and algorithms. *International Journal in Human-Computer Studies*, 59(1–2):157–183, 2003.

[504] S. L. Oviatt, C. Darves, and R. Coulston. Toward adaptive conversational interfaces: Modeling speech convergence with animated personas. *ACM Transactions on Computer-Human Interaction*, 11(3):300–328, 2004.

[505] M. M. Bradley P. J. Lang. Emotion and the motivational brain. *Biological Psychology*, 84(3):437–450, 2010.

[506] F. Pallavicini, D. Algeri, C. Repetto, A. Gorini, and G. Riva. Biofeedback, virtual reality and mobile phones in the treatment of Generalized Anxiety Disorder (GAD): A phase-2 controlled clinical trial. *Journal of CyberTherapy & Rehabilitation*, 2(4):315–327, 2009.

[507] J. S. Pan, F. R. McInnes, and M. A. Jack. Bound for Minkowski metric or quadratic metric applied to VQ codeword search. *IEE Proceedings – Vision, Image and Signal Processing*, 143(1):67–71, 1996.

[508] A. Pantelopoulos and N. G. Bourbakis. A survey on wearable sensor-based systems

for health monitoring and prognosis. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, 40(1):1–12, 2010.

[509] M. Pantic. *Facial expression analysis by computational intelligence techniques*. PhD thesis, Faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, 2001.

[510] M. Pantic and I. Patras. Dynamics of facial expression: Recognition of facial actions and their temporal segments from face profile image sequences. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 36(2):433–449, 2006.

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

[512] S. Paulmann, D. Titone, and M. D. Pell. How emotional prosody guides your way: Evidence from eye movements. *Speech Communication*, [in press].

[513] T. Pawar, N. S. Anantakrishnan, S. Chaudhuri, T. D. Pawar, and S. Duttagupta. Impact of ambulation in wearable-ECG. *Annals of Biomedical Engineering*, 36(9):1547–1557, 2008.

[514] M. Perc. Nonlinear time series analysis of the human electrocardiogram. *European Journal of Physics*, 26(5):757–768, 2005.

[515] P. Petta, C. Pelachaud, and R. Cowie. *Emotion-oriented systems: The Humaine handbook*. Cognitive Technologies. Berlin/Heidelberg, Germany: Springer-Verlag, 2011.

[516] H. Pham. *Springer handbook of engineering statistics*. London, UK: Springer-Verlag London Limited, 2006.

[517] L. Phan and S. Barriga. Nico Frijda: Opening the Pandora's box of sciences. *the free mind*, 1(2):22–29, 2006.

[518] R. Picard. Affective computing for HCI. In *Human-Computer Interaction: Ergonomics and User Interfaces (Proceedings of HCI International '99)*, volume 1, pages 829–833, Munich, Germany, August 22–27 1999. Mahwah, NJ, USA: Lawrence Erlbaum Associates, Inc.

[519] R. R. Picard and R. D. Cook. Cross-validation of regression models. *Journal of the American Statistical Association*, 79(387):575–583, 1984.

[520] R. W. Picard. Affective computing. Technical Report 321, M.I.T. Media Laboratory Perceptual Computing Section, Cambridge, MA, USA, 1995.

[521] R. W. Picard. *Affective Computing*. Boston, MA, USA: MIT Press, 1997.

[522] R. W. Picard. Toward computers that recognize and respond to user emotion. *IBM Systems Journal*, 39(3&4):705–719, 2000.

[523] R. W. Picard. and J. Scheirer. The galvactivator: A glove that senses and communicates skin conductivity – usability evaluation and interface design: Cognitive engineering, intelligent agents and virtual reality. In M. J. Smith, G. Salvendy, D. Harris, and R. J. Koubek, editors, *Proceedings of the 9th International Conference on Human-Computer Interaction*, volume 1, pages 1538–1542, New Orleans, Louisiana, USA, 5–10 August 2001.

[524] R. W. Picard, E. Vyzas, and J. Healey. Toward machine emotional intelligence: Analysis of affective physiological state. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(10):1175–1191, 2001.

[525] K. D. Pimple. Computing ethics: Surrounded by machines. *Communications of the ACM*, 54(3):29–31, 2011.

[526] J. Pittermann, A. Pittermann, and W. Minker. *Handling emotions in human-computer dialogues*. Dordrecht, The Netherlands: Springer Science+Business Media B.V., 2010.

[527] M. Pivec. *Affective and emotional aspects of Human-Computer Interaction: Game-based and innovative learning approaches*, volume 1 of *The Future of Learning*. Amsterdam, Netherlands: IOS Press, 2006.

[528] R. Plamondon and S. N. Srihari. Online and offline handwriting recognition: A comprehensive survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(1):63–84, 2000.

[529] K. Poels and S. Dewitte. How to capture the heart? Reviewing 20 years of emotion measurement in advertising. *Journal of Advertising Research*, 46(1):18–37, 2006.

[530] M.-Z. Poh, N. C. Swenson, and R. W. Picard. A wearable sensor for unobtrusive, long-term assessment of electrodermal activity. *IEEE Transactions on Biomedical Engineering*, 14(2):1243–1252, 2010.

[531] M. J. Power and T. Dalgleish. *Cognition and emotion: From order to disorder*. Hove, East Sussex, UK: Psychology Press / Taylor & Francis Group, 2nd edition, 2008.

[532] J. Preece. Empathic communities: Reaching out across the Web. *Interactions*, 5(2):32–43, 1998.

[533] H. Prendinger and M. Ishizuka. *Life-like characters: Tools, affective functions, and applications*. Cognitive Technologies. Berlin/Heidelberg, Germany: Springer-Verlag, 2010.

[534] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical recipes:*

The art of scientific computing. Cambridge, UK: Cambridge University Press, 3rd edition, 2007.

[535] J. J. Prinz. *Gut Reactions: A Perceptual Theory of Emotion*. Philosophy of Mind series. New York, NY, USA: Oxford University Press, Inc., 2004.

[536] W. F. Prokasy and D. C. Raskin. *Electrodermal activity in psychological research*. New York, NY, USA: Academic Press, Inc., 1973.

[537] L. Rabiner, M. Cheng, A. Rosenberg, and C. McGonegal. A comparative performance study of several pitch detection algorithms. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 24(5):399–418, 1976.

[538] P. Rainville, A. Bechara, N. Naqvi, and A. R. Damasio. Basic emotions are associated with distinct patterns of cardiorespiratory activity. *International Journal of Psychophysiology*, 61(1):5–18, 2006.

[539] R. J. E. Rajae-Joordens. *The effects of coloured light on valence and arousal*, volume 12 of *Philips Research Book Series*, chapter 5, pages 65–84. Dordrecht, The Netherlands: Springer Science+Business Media B.V., 2011.

[540] C. Ramos, J. C. Augusto, and D. Shapiro. Ambient Intelligence – The next step for Artificial Intelligence. *IEEE Intelligent Systems*, 23(2):15–18, 2008.

[541] P. Rani, C. Liu, N. Sarkar, and E. Vanman. An empirical study of machine learning techniques for affect recognition in human-robot interaction. *Pattern Analysis & Applications*, 9(1):58–69, 2006.

[542] P. Rani, N. Sarkar, and J. Adams. Anxiety-based affective communication for implicit human-machine interaction. *Advanced Engineering Informatics*, 21(3):323–334, 2007.

[543] P. Rani, N. Sarkar, C. Smith, and L. Kirby. Anxiety detecting robotic systems – towards implicit human-robot collaboration. *Robotica*, 22(1):85–95, 2004.

[544] P. Rani, N. Sarkar, C. A. Smith, and J. Adams. Affective communication for implicit human-machine interaction. In *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*, volume 5, pages 4896–4903, Washington, D.C., USA, October 5–8 2003. Piscataway, NJ, USA: IEEE, Inc.

[545] M. Rauterberg. An empirical comparison of menu-selection (CUI) and desktop (GUI) computer programs carried out by beginners and experts. *Behaviour & Information Technology*, 11(4):227–236, 1992.

[546] R. L. Ray and H. D. Kimmel. Utilization of psychophysiological indices in behavioral assessment: Some methodological issues. *Journal of Behavioral Assessment*, 1(2):107–122, 1979.

[547] S. L. Ray. Evolution of posttraumatic stress disorder and future directions. *Archives of Psychiatric Nursing*, 22(4):217–225, 2008.

[548] M. B. I. Reaz, M. S. Hussain, and F. Mohd-Yasin. Techniques of EMG signal analysis: detection, processing, classification and applications. *Biological Procedures Online*, 8(1):11–35, 2006.

[549] B. Reeves and C. Nass. *The media equation: How People Treat Computers, Television, and New Media Like Real People and Places*. New York, N.Y., U.S.A.: Cambridge University Press, 1996.

[550] T. Regier and P. Kay. Language, thought, and color: Whorf was half right. *Trends in Cognitive Sciences*, 13(10):411–454, 2009.

[551] A. C. Rencher. *Methods of Multivariate Analysis*. Wiley Series in Probability and Statistics. New York, NY, USA: John Wiley & Sons, Inc., 2nd edition, 2002.

[552] S. O. Rice. Mathematical analysis of random noise. ATI 162 535 Monograph B-1589, Bell Telephone Labs, Inc., New York, NY, USA, January 1952.

[553] D. Richter, C. Dietzel, and U. Kunzmann. Age differences in emotion recognition: The task matters. *Journal of Gerontology: Psychological Sciences*, 66B(1):48–55, 2010.

[554] S.W. Rienstra and A. Hirschberg. *An Introduction to Acoustics*. Eindhoven, The Netherlands: Eindhoven University of Technology, 5th edition, 2009. Extended and revised edition of IWDE 92-06 (TU/e).

[555] I. Rivals and L. Personnaz. On cross validation for model selection. *Neural Computation*, 11(4):863–870, 1998.

[556] N. A. Roberts and R. W. Levenson. Subjective, behavioral, and physiological reactivity to ethnically matched and ethnically mismatched film clips. *Emotion*, 6(4):635–646, 2006.

[557] N. A. Roberts, R. W. Levenson, and J. J. Gross. Cardiovascular costs of emotion suppression cross ethnic lines. *International Journal of Psychophysiology*, 70(1):82–87, 2008.

[558] P. Robinson and R. el Kaliouby. *Computation of emotions in man and machines*. Philosophical Transactions of the Royal Society series B – Biological Sciences, 364(1535), 3439–3604. London, UK: The Royal Society, 2009.

[559] M. Rogers. The role of criminal profiling in the computer forensics process. *Computers & Security*, 22(4):292–298, 2003.

[560] L. J. M. Rothkrantz, P. Wiggers, J.-W. A. van Wees, and R. J. van Vark. Voice stress

analysis. *Lecture Notes in Computer Science (Text, Speech and Dialogue)*, 3206:449–456, 2004.

[561] S. Rothman. *Physiology and biochemistry of the skin*. Chicago, IL, USA: The University of Chicago Press, 1954.

[562] D. C. Rubin, D. Berntsen, and M. Klindt Bohni. A memory-based model of posttraumatic stress disorder: Evaluating basic assumptions underlying the PTSD diagnosis. *Psychological Review*, 115(4):985–1011, 2008.

[563] N. Rüscha, P. W. Corrigan, K. Powell, A. Rajah, M. Olschewski, S. Wilkniss, and K. Batia. A stress-coping model of mental illness stigma: II. emotional stress responses, coping behavior and outcome. *Schizophrenia Research*, 110(1–3):65–71, 2009.

[564] N. Rüscha, P. W. Corrigan, A. Wassel, P. Michaels, M. Olschewski, S. Wilkniss, and K. Batia. A stress-coping model of mental illness stigma: I. predictors of cognitive stress appraisal. *Schizophrenia Research*, 110(1–3):59–64, 2009.

[565] J. A. Russel and L. F. Barrett. Core affect, prototypical emotional episodes, and other things called emotion: Dissecting the elephant. *Journal of Personality and Social Psychology*, 26(5):805–819, 1999.

[566] J. A. Russell. A circumplex model of affect. *Journal of Personality and Social Psychology*, 39(6):1161–1178, 1980.

[567] J. A. Russell. Core affect and the psychological construction of emotion. *Psychological Review*, 110(1):145–172, 2003.

[568] S. J. Russell and P. Norvig. *Artificial Intelligence: A Modern Approach*. Prentice Hall series in Artificial Intelligence. Upper Saddle River, NJ, USA: Pearson Education, Inc., 3rd edition, 2010.

[569] T. M. Rutkowski, T. Tanaka, A. Cichocki, D. Erickson, J. Cao, and D. P. Mandic. Interactive component extraction from fEEG, fNIRS and peripheral biosignals for affective brain-machine interfacing paradigms. *Computers in Human Behavior*, 27(5):1512–1518, 2011.

[570] K. Rymarczyk, C. Biele, A. Grabowska, and H. Majczynski. EMG activity in response to static and dynamic facial expressions. *International Journal of Psychophysiology*, 79(2):330–333, 2011.

[571] A. Sakurai, M. Nakamura, and J. Nakamura. Self-organizing map analysis of sensor networks for human movement tracking. *Sensors and Actuators A: Physical*, 166(1):141–148, 2011.

[572] J. Sánchez-Meca, A. I. Rosa-Alcázar, F. Marín-Martínez, and A. Gómez-Conesa. Psy-

chological treatment of panic disorder with or without agoraphobia: A meta-analysis. *Clinical Psychology Review*, 30(1):37–50, 2010.

[573] D. Sander and K. R. Scherer. *The Oxford companion to emotion and affective sciences*. Series in Affective Science. Oxford, NY, USA: Oxford University Press Inc., 1st edition, 2009.

[574] A. Savran, B. Sankur, and M. T. Bilge. Comparative evaluation of 3D versus 2D modality for automatic detection of facial action units. *Pattern Recognition*, [in press].

[575] S. Schachter and J. E. Singer. Cognitive, social, and physiological determinants of emotional state. *Psychological Review*, 69(5):379–399, 1962.

[576] J. Scheirer, R. Fernandez, J. Klein, and R. W. Picard. Frustrating the user on purpose: A step toward building an affective computer. *Interacting with Computers*, 14(2):93–118, 2002.

[577] A. M. Schell, M. E. Dawson, and D. L. Filion. Psychophysiological correlates of electrodermal activity. *Psychophysiology*, 25(6):619–632, 1988.

[578] K. R. Scherer. Studying the emotion-antecedent appraisal process: An expert system approach. *Cognition & Emotion*, 7(34):325–355, 1993.

[579] K. R. Scherer. Vocal communication of emotion: A review of research paradigms. *Speech Communication*, 40(1–2):227–256, 2003.

[580] K. R. Scherer. The dynamic architecture of emotion: Evidence for the component process model. *Cognition & Emotion*, 23(7):1307–1351, 2009. individual differences.

[581] K. R. Scherer, R. P. Abeles, and C. S. Fischer. *Human aggression and conflict. Interdisciplinary perspectives*. Englewood Cliffs, NJ, USA: Prentice-Hall, Inc., 1975.

[582] K. R. Scherer, T. Bänziger, and E. B. Roesch. *Blueprint for Affective Computing: A sourcebook*. Series in Affective Science. New York, NY, USA: Oxford University Press, Inc., 2010.

[583] K. R. Scherer and H. Giles. *Social markers in speech*. European Studies in Social Psychology. Cambridge, Great Britain: Cambridge University Press, 1979.

[584] K. R. Scherer, H. G. Wallbott, and A. B Summerfield. *Experiencing emotion: A cross-cultural study*. European Monographs in Social Psychology. Cambridge, UK: Cambridge University Press, 2011.

[585] N. Schneiderman, S. M. Weiss, and P. G. Kaufmann. *Handbook of research methods in cardiovasular behavioral medicine*. The Plenum series in Behavioral Psychophysiology and Medicine. New York, NY, USA: Plenum Press, 1989.

[586] B. Schölkopf and A. J. Smola. *Learning with kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. Adaptive Computation and Machine Learning. Cambridge, MA, USA: The MIT Press, 2002.

[587] L. Schomaker, J. Nijtmans, A. Camurri, F. Lavagetto, P. Morasso, C. Benoît, T. Guiard-Marigny, B. Le Goff, J. Robert-Ribes, A. Adjoudani, I. Defée, S. Münch, K. Hartung, and J. Blauert. A taxonomy of multimodal interaction in the human information processing system. Technical report, Esprit Project 8579 MIAMI – WP 1, 1995.

[588] M. Schröder. *Speech and Emotion Research: An Overview of Research Frameworks and a Dimensional Approach to Emotional Speech Synthesis*. PhD thesis, Philosophischen Fakultäten der Universität des Saarlandes, 2004.

[589] J. L. H. Schuler and W. H. O'Brien. Cardiovascular recovery from stress and hypertension factors: A meta-analytic view. *Psychophysiology*, 34(6):649–659, 1997.

[590] B. Schuller, A. Batliner, S. Steidl, and D. Seppi. Recognising realistic emotions and affect in speech: State of the art and lessons learnt from the first challenge. *Speech Communication*, 53(9–10):1062–1087, 2011.

[591] H. T. Schupp, B. N. Cuthbert, M. M. Bradley, C. H. Hillman, A. O. Hamm, and P. J. Lang. Brain processes in emotional perception: Motivated attention. *Cognition & Emotion*, 18(5):593–611, 2004.

[592] G. E. Schwartz, P. L. Fair, P. Salt, M. R. Mandel, and G. L. Klerman. Facial muscle patterning to affective imagery in depressed and nondepressed subjects. *Science*, 192(4238):489–491, 1976.

[593] N. Schwarz. Emotion, cognition, and decision making. *Cognition & Emotion*, 14(4):433–440, 2000.

[594] A. Sears and J. A. Jacko. *The Human-Computer Interaction handbook: Fundamentals, evolving technologies and emerging applications*. Human Factors and Ergonomics. New York, NY, USA: Lawrence Erlbaum Associates / Taylor & Francis Group, LLC, 2nd edition, 2008.

[595] N. Sebe. Multimodal interfaces: Challenges and perspectives. *Journal of Ambient Intelligence and Smart Environments*, 1(1):23–30, 2009.

[596] G. A. F. Seber and C. J. Wild. *Nonlinear Regression*. Hoboken, NJ, USA: John Wiley & Sons, Inc., reproduction edition, 1989/2003.

[597] D. Sekulovski, I. Vogels, R. Clout, and M. Perz. Changing color over time. *Lecture Notes in Computer Science (Ergonomics and Health Aspects)*, 6779:218–225, 2011.

[598] M. E. P. Seligman and M. Csikszentmihalyi. Positive psychology: An introduction. *American Psychologist*, 55(1):5–14, 2000.

[599] P. Seuntiens, I. Vogels, and A. van Keersop. Visual experience of 3D-TV with pixelated Ambilight. In *Proceedings of PRESENCE 2007: The 10th Annual International Workshop on Presence*, pages 339–344, Barcelona, Spain, October 25–27 2007. Barcelona, Spain: Starlab Barcelona S.L / International Society for Presence Research.

[600] N. Sharkey and A. Sharkey. The crying shame of robot nannies: An ethical appraisal. *Interaction Studies*, 11(2):161–190, 2010.

[601] T. Sharot and A. P. Yonelinas. Differential time-dependent effects of emotion on recollective experience and memory for contextual information. *Cognition*, 106(1):538–547, 2008.

[602] J. Shawe-Taylor and S. Sun. A review of optimization methodologies in support vector machines. *Neurocomputing*, [in press].

[603] B.-J. Shen, L. R. Stroud, and R. Niaura. Ethnic differences in cardiovascular responses to laboratory stress: A comparison between Asian and white Americans. *International Journal of Behavioral Medicine*, 11(3):181–186, 2004.

[604] L. Shen, V. Callaghan, and R. Shen. Affective e-learning in residential and pervasive computing environments. *Information Systems Frontiers*, 10(4):461–472, 2008.

[605] S. Shepperd, H. Doll, J. Broad, J. Gladman, S. Iliffe, P. Langhorne, S. Richards, F. Martin, and R. Harris. Early discharge hospital at home. *Cochrane Database of Systematic Reviews*, 2009(1):Article No. CD000356, 2009.

[606] A. Sherwood, M. T. Allen, J. Fahrenberg, R. M. Kelsey, W. R. Lovallo, and L. J. van Doornen. Methodological guidelines for impedance cardiography. *Psychophysiology*, 27(1):1–23, 1990.

[607] T. Shimamura and H. Kobayashi. Weighted autocorrelation for pitch extraction of noisy speech. *IEEE Transactions on Speech and Audio Processing*, 9(7):727–730, 2001.

[608] E. H. Shortliffe. *Computer-based medical consultations: MYCIN*, volume 2 of *Artificial Intelligence Series*. New York, NY, USA: Elsevier Science Ltd., 1976.

[609] B. Siciliano and O. Khatib. *Springer Handbook of Robotics*. Springer Handbooks. Berlin/Heidelberg, Germany: Springer-Verlag, 2008.

[610] J. Sifakis. A vision for computer science – the system perspective. *Central European Journal of Computer Science*, 1(1):108–116, 2011.

[611] H. A. Simon. Motivational and emotional controls of cognition. *Psychological Review*, 74(1):29–39, 1967.

[612] T. Singer and C. Lamm. The social neuroscience of empathy. *Annals of the New York Academy of Sciences*, 1156(1):81–96, 2009.

[613] R. Sinha and O. A. Parsons. Multivariate response patterning of fear. *Cognition and Emotion*, 10(2):173–198, 1996.

[614] B. F. Skinner. *Beyond freedom and dignity*. New York, NY, USA: Alfred A. Knopf Inc., 1971.

[615] M. Slater, C. Guger, G. Edlinger, R. Leeb, G. Pfurtscheller, A. Antley, M. Garau, A. Brogni, and D. Friedman. Analysis of physiological responses to a social situation in an immersive virtual environment. *Presence: Teleoperators and Virtual Environments*, 15(5):553–569, 2006.

[616] M. Slater, P. Khanna, J. Mortensen, and I. Yu. Visual realism enhances realistic response in an immersive virtual environment. *IEEE Computer Graphics and Applications*, 29(3):76–84, 2009.

[617] M. Smith. *Literacy and Augmentative and Alternative Communication*. Augmentative and Alternative Communication Perspectives Series. Burlington, MA, USA: Elsevier Academic Press, 2005.

[618] T. Sobol-Shikler and P. Robinson. Classification of complex information: Inference of co-occurring affective states from their expressions in speech. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(7):1284–1297, 2010.

[619] E. D. Sontag. *Mathematical control theory: Deterministic finite dimensional systems*, volume 6 of *Texts in Applied Mathematics*. New York, NY, USA: Springer-Verlag New York, Inc., 2nd edition, 1998.

[620] L. Sörnmo and P. Laguna. *Bioelectrical Signal Processing in Cardiac and Neurological Applications*. Burlington, MA, USA: Elsevier Academic Press, 2005.

[621] S. Stankov, V. Glavinić, and M. Rosić. *Intelligent Tutoring Systems in E-Learning Environments: Design, Implementation and Evaluation*. Premier Reference Source. Hershey, PA, USA: Information Science Reference / IGI Global, 2011.

[622] L. Stead, P. Goulev, C. Evans, and E. Mamdani. The emotional wardrobe. *Personal and Ubiquitous Computing*, 8(3–4):282–290, 2004.

[623] G. Stemmler, T. Aue, and J. Wacker. Anger and fear: Separable effects of emotion and motivational direction on somatovisceral responses. *International Journal of Psychophysiology*, 66(2):141–153, 2007.

[624] G. Stemmler and J. Wacker. Personality, emotion, and individual differences in physiological responses. *Biological Psychology*, 83(3):541–551, 2010.

[625] A. Steptoe and M. Marmot. Impaired cardiovascular recovery following stress predicts 3-year increases in blood pressure. *Journal of Hypertension*, 23(3):529–536, 2005.

[626] R. M. Stern, W. J. Ray, and K. S. Quigley. *Psychophysiological Recording*. New York, NY, USA: Oxford University Press, 2nd edition, 2001.

[627] R. A. Sternbach and B. Tursky. Ethnic differences among housewives in psychophysical and skin potential responses to electric shock. *Psychophysiology*, 1(3):241–246, 1965.

[628] M. Stone. Cross-validation choice and assessment of statistical predictions. *Journal of the Royal Statistical Society. Series B (Methodological)*, 36(2):111–147, 1974.

[629] M. Stone. An asymptotic equivalence of choice of model by cross-validation and Akaike's criterion. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1):44–47, 1977.

[630] M. Stone. Asymptotics for and against cross-validation. *Biometrika*, 64(1):29–35, 1977.

[631] K. T. Strongman. *The psychology of emotion: From everyday life to theory*. Chichester, West Sussex, England: John Wiley & Sons Ltd, 5th edition, 2003.

[632] A. Stuiver and L. J. M. (Ben) Mulder. On artefact-free real-time computation of cardiovascular measures. In *Proceedings of the IEEE 3rd International Conference on Affective Computing and Intelligent Interaction, ACII*, volume 1, pages 716–721, Amsterdam, The Netherlands, 10–12 September 2009. IEEE Press.

[633] N. Suwa and I. Yamashita. *Psychophysiological studies of emotion and mental disorders*. Tokyo, Japan: IGAKU SHOIN LTD., 1974.

[634] K. Takahashi. Remarks on emotion recognition from bio-potential signals. In *Proceedings of the 2nd International Conference on Autonomous Robots and Agents*, pages 186–191, Palmerston North, New Zealand, December 13–15 2004.

[635] K. Takahashi. Remarks on SVM–based emotion recognition from multi-modal bio-potential signals. In *Proceedings of the 13th IEEE International Workshop on Robot and Human Interactive Communication*, pages 95–100, Kurashiki, Okayama Japan, September 20–22 2004. Piscataway, NJ, USA: IEEE, Inc.

[636] K. Takahashi and A. Tsukaguchi. Remarks on emotion recognition from multi-modal bio-potential signals. In *IEEE International Conference on Systems, Man and Cybernetics*, volume 2, pages 1654–1659, Washington, D.C., USA, October 5–8 2003. Piscataway, NJ, USA: IEEE, Inc.

[637] D. S. Tan and A. Nijholt. *Brain-Computer Interfaces: Applying our minds to human-*

computer interaction. Human-Computer Interaction Series. London, Great Britain: Springer-Verlag London Limited, 2010.

[638] J. Tan. *Medical informatics: Concepts, methodologies, tools, and applications*. Premier Reference Source. Hershey, PA, USA: Information Science Reference / IGI Global, 2009.

[639] J. Tao and T. Tan. *Affective Information Processing*. London, UK: Springer-Verlag, 2009.

[640] V. C. Tartter. Happy talk: Perceptual and acoustic effects of smiling on speech. *Perception & Psychophysics*, 27(1):24–27, 1980.

[641] Task Force of the European Society of Cardiology the North American Society of Pacing Electrophysiology. Heart rate variability: Standards of measurement, physiological interpretation, and clinical use. *Circulation*, 93(5):1043–1065, 1996.

[642] L. G. Tassinary and B. S. Marshall and. Microcomputers in social psychophysiological research: An overview. *Behavior Research Methods, Instruments, & Computers*, 17(5):532–536, 1985.

[643] L. G. Tassinary and J. T. Cacioppo. Unobservable facial actions and emotion. *Psychological Science*, 3(1):28–33, 1992.

[644] A. Tawari and M. M. Trivedi. Speech emotion analysis: Exploring the role of context. *IEEE Transactions on MultiMedia*, 12(6):502–509, 2010.

[645] S. E. Taylor and A. L. Stanton. Coping resources, coping processes, and mental health. *Annual Review of Clinical Psychology*, 3:377–401, 2007.

[646] A. Teeters, R. el Kaliouby, and R. Picard. Self-Cam: Feedback from what would be your social partner. In *Proceedings of SIGGRAPH: The 33rd International Conference on Computer Graphics and Interactive Techniques*, page Article no. 138, Boston, MA, USA, July 30 – August 03 2006. New York, NY, USA: ACM.

[647] R. E. Thayer. *The biopsychology of mood and arousal*. New York, NY, USA: Oxford University Press, Inc., 1989.

[648] S. Theodoridis and K. Koutroumbas. *Pattern Recognition*. Burlington, MA, USA: Academic Press / Elsevier Inc., 4th edition, 2009.

[649] S. Thiemjarus and G.-Z. Yang. *Context-Aware Sensing*, chapter 9, pages 287–331. London, UK: Springer-Verlag London Limited, 2006.

[650] J. K. Thompson, J. D. Haber, and B. H. Tearnan. Generalization of frontalis electromyographic feedback to adjacent muscle groups: A critical review. *Psychosomatic Medicine*, 43(1):19–24, 1981.

[651] V. A. Thurmond. The point of triangulation. *Journal of Nursing Scholarship*, 33(3):253–258, 2001.

[652] Y.-L. Tian, T. Kanade, and J. F. Cohn. *Facial expression analysis*, chapter 11, pages 247–275. New York, NY, USA: Springer Science+Business Media, Inc., 2005.

[653] I. R. Titze and E. J. Hunter. A two-dimensional biomechanical model of vocal fold posturing. *The Journal of the Acoustical Society of America*, 121(4):2254–2260, 2007.

[654] N. Tractinsky. Tools over solutions? comments on interacting with computers special issue on affective computing. *Interacting with Computers*, 16(4):751–757, 2004.

[655] P. Trahanias and E. Skordalakis. Syntactic pattern recognition of the ECG. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(7):648–657, 1990.

[656] R. Trappl, P. Petta, and S. Payr. *Emotions in Humans and Artifacts*. Cambridge, MA, USA: The MIT Press, 2003.

[657] M. Trimmel, C. Goger, E.-M. Vouk, M. Kritz, A. Klaus, and E. Groll-Knapp. Heart rate of pilots during long haul flights indicate low arousal states. *Psychophysiology*, 42(S1):S26, 2005.

[658] W. L. Tse and W. L. Chan. A distributed sensor network for measurement of human thermal comfort feelings. *Sensors and Actuators A: Physical*, 144(2):394–402, 2008.

[659] J. R. Turner, A. Sherwood, and K. C. Light. *Individual differences in cardiovascular response to stress*. Perspectives on individual differences. New York, NY, USA: Plenum Press, 1992.

[660] B. Tursky and R. A. Sternbach. Further physiological correlates of ethnic differences in responses to shock. *Psychophysiology*, 4(1):67–74, 1967.

[661] B. N. Uchino, D. Uno, J. Holt-Lunstad, and J. B. Flinders. Age-related differences in cardiovascular reactivity during acute psychological stress in men and women. *Journals of Gerontology: Psychological Sciences*, 54B(6):P339–P346, 1999.

[662] J. Vallverdú and D. Casacuberta. *Handbook of research on synthetic emotions and sociable robotics: New applications in Affective Computing and Artificial Intelligence*. Premier Reference Source. Hershey, PA, USA: Information Science Reference / IGI Global, 2009.

[663] A. van Boxtel. Optimal signal bandwidth for the recording of surface emg activity of facial, jaw, oral, and neck muscles. *Psychophysiology*, 38(1):22–34, 2001.

[664] A. van Boxtel. Facial EMG as a tool for inferring affective states. In A. J. Spink, F. Grieco, O. E. Krips, L. W. S. Loijens, L. P. J. J. Noldus, and P. H. Zimmerman, editors, *Proceedings of Measuring Behavior 2010: 7th International Conference on Methods*

and Techniques in Behavioral Research, pages 104–108, Eindhoven, The Netherlands, August 24–27 2010. Noldus Information Technology.

[665] A. van Boxtel, P. Goudswaard, and L. R. B. Schomaker. Recording methods for the frontalis surface EMG. *Psychophysiology*, 20:475, 1983.

[666] A. van Boxtel, P. Goudswaard, and L. R. B. Schomaker. Amplitude and bandwidth of the frontalis surface EMG: Effects of electrode parameters. *Psychophysiology*, 21(6):699–707, 1984.

[667] A. van Boxtel and L. R. B. Schomaker. Motor unit firing rate during static contraction indicated by the surface EMG power spectrum. *IEEE Transactions on Biomedical Engineering*, BME-30(9):601–608, 1983.

[668] A. H. van Bunningen, L. Feng, and P. M. G. Apers. Context for ubiquitous data management. In *IEEE Proceedings of the International Workshop on Ubiquitous Data Management (UDM2005)*, pages 17–24, Tokyo, Japan, April 4 2005. Los Alamitos, CA, USA: IEEE Computer Society.

[669] A. H. van Bunningen, L. Feng, and P. M. G. Apers. A context-aware preference model for database querying in an Ambient Intelligent environment. *Lecture Notes in Computer Science (Database and Expert Systems Applications)*, 4080:33–43, 2006.

[670] E. L. van den Broek. Emotional Prosody Measurement (EPM): A voice-based evaluation method for psychological therapy effectiveness. *Studies in Health Technology and Informatics (Medical and Care Compunetics)*, 103:118–125, 2004.

[671] E. L. van den Broek. Empathic Agent Technology (EAT). In L. Johnson, D. Richards, E. Sklar, and U. Wilensky, editors, *Proceedings of the AAMAS-05 Agent-Based Systems for Human Learning (ABSHL) workshop*, pages 59–67, Utrecht - The Netherlands, 2005.

[672] E. L. van den Broek. Robot nannies: Future or fiction? *Interaction Studies*, 11(2):274–282, 2010.

[673] E. L. van den Broek. Ubiquitous emotion-aware computing. *Personal and Ubiquitous Computing*, 15:[in press], 2011.

[674] E. L. van den Broek, J. H. Janssen, and J. H. D. M. Westerink. Guidelines for Affective Signal Processing (ASP): From lab to life. In *Proceedings of the IEEE 3rd International Conference on Affective Computing and Intelligent Interaction, ACII*, volume 1, pages 704–709, Amsterdam, The Netherlands, 10–12 September 2009. IEEE Press.

[675] E. L. van den Broek, V. Lisý, J. H. Janssen, J. H. D. M. Westerink, M. H. Schut, and K. Tuinenbreijer. *Affective Man-Machine Interface: Unveiling human emotions*

through biosignals, volume 52 of *Communications in Computer and Information Science*, pages 21–47. Berlin/Heidelberg, Germany: Springer-Verlag, 2010.

[676] E. L. van den Broek, M. H. Schut, J. H. D. M. Westerink, and K. Tuinenbreijer. Unobtrusive Sensing of Emotions (USE). *Journal of Ambient Intelligence and Smart Environments*, 1(3):287–299, 2009.

[677] E. L. van den Broek, F. van der Sluis, and T. Dijkstra. *Telling the story and re-living the past: How speech analysis can reveal emotions in post-traumatic stress disorder (PTSD) patients*, volume 12 of *Philips Research Book Series*, chapter 10, pages 153–180. Dordrecht, The Netherlands: Springer Science+Business Media B.V., 2011.

[678] E. L. van den Broek, F. van der Sluis, and Th. E. Schouten. User-centered digital preservation of multimedia. *ERCIM (European Research Consortium for Informatics and Mathematics) News*, 80:45–47, 2010.

[679] E. L. van den Broek and Joyce H. D. M. Westerink. Considerations for emotion-aware consumer products. *Applied Ergonomics*, 40(6):1055–1064, 2009.

[680] E. L. van den Broek et al. Prerequisites for Affective Signal Processing (ASP) – Parts I–V. In A. Fred, J. Filipe, and H. Gamboa, editors, *BioSTEC 2009/2010/2011: Proceedings of the International Joint Conference on Biomedical Engineering Systems and Technologies*, pages –, Porto, Portugal / Valencia, Spain / Rome, Italy, 2009/2010/2011. INSTICC – Institute for Systems and Technologies of Information, Control and Communication.

[681] M. D. van der Zwaag, J. H. D. M. Westerink, and E. L. van den Broek. Emotional and psychophysiological responses to tempo, mode, and percussiveness. *Musicae Scientiae*, 15(2):250–269, 2011.

[682] A. van Drunen, E. L. van den Broek, A. J. Spink, and T. Heffelaar. Exploring workload and attention measurements with uLog mouse data. *Behavior Research Methods*, 41(3):868–875, 2009.

[683] M. van Gerven, J. Farquhar, R. Schaefer, R. Vlek, J. Geuze, A. Nijholt, N. Ramsay, P. Haselager, L. Vuurpijl, S. Gielen, and P. Desain. The brain-computer interface cycle. *Journal of Neural Engineering*, 6(4):1–10, 2009.

[684] H. van Heerde, N. Anciaux, L. Feng, and P. M. G. Apers. Balancing smartness and privacy for the Ambient Intelligence. *Lecture Notes in Computer Science (Smart Sensing and Context)*, 4272:255–258, 2006.

[685] A. van Minnen, I. Wessel, T. Dijkstra, and K. Roelofs. Changes in PTSD patients' narratives during prolonged exposure therapy: A replication and extension. *Journal of Traumatic Stress*, 15(3):255–258, 2002.

[686] H. C. van Vugt, E. A. Konijn, J. F. Hoorn, I. Keur, and A. Eliëns. Realism is not all! user engagement with task-related interface characters. *Interacting with Computers*, 19(2):267–280, 2007.

[687] H. C. van Vugt, J. F. van Hoorn, E. A. Konijn, and A. de Bie Dimitriadou. Affective affordances: Improving interface character engagement through interaction. *International Journal of Human-Computer Studies*, 64(9):874–888, 2006.

[688] V. Vapnik. *Estimation of dependences based on empirical data*. Information Science & Statistics. New York, NY, USA: Springer Science+Business Media, Inc., 2nd edition, 1982/2006.

[689] V. N. Vapnik. *Statistical Learning Theory*. Adaptive and Learning Systems for Signal Processing, Communications, and Control. New York, NY, USA: John Wiley & Sons, Inc., 1998.

[690] V. N. Vapnik. An overview of statistical learning theory. *IEEE Transactions on Neural Networks*, 10(5):988–999, 1999.

[691] V. N. Vapnik. *The Nature of Statistical Learning Theory*. Statistics for Engineering and Information Science. New York, NY, USA: Springer-Verlag New York, Inc., 2nd edition, 2000.

[692] R. Veenhoven. Healthy happiness: Effects of happiness on physical health and the consequences for preventive health care. *Journal of Happiness Studies*, 9(3):449–469, 2008.

[693] V. Venkatesh. Determinants of perceived ease of use: Integrating control, intrinsic motivation, and emotion into the technology acceptance model. *Information Systems Research*, 11(4):342–365, 2000.

[694] R. Ventura and C. Pinto-Ferreira. Responding efficiently to relevant stimuli using an emotion-based agent architecture. *Neurocomputing*, 72(13–15):2923–2930, 2009.

[695] R. Vertegaal. Attentive user interfaces. *Communications of the ACM*, 46(3):30–33, 2003.

[696] D. Ververidis and C. Kotropoulos. Emotional speech recognition: Resources, features, and methods. *Speech Communication*, 48(9):1162–1181, 2006.

[697] O. Villon and C. L. Lisetti. A user model of psycho-physiological measure of emotion. *Lecture Notes in Computer Science (User Modelling)*, 4511:319–323, 2007.

[698] P. Vincent and Y. Bengio. *K-local hyperplane and convex distance nearest neighbor algorithms*, volume 14 II (Part IV Algorithms and Architectures) of *NIPS Proceedings*

Series, pages 985–992. Cambridge, MA, USA: Massachusetts Institute of Technology, 2002.

[699] A. Vinciarelli. Capturing order in social interactions. *IEEE Signal Processing Magazine*, 26(12):133–137, 2009.

[700] A. J. H. M. Vingerhoets and L. R. B. Schomaker. *Emotional fainting: Its physiological and psychological aspects*, volume 12 of *The Series in Clinical and Community Psychology*, chapter 13 (Part III: Physiological aspects of stress), pages 181–198. New York, NY, USA: Hemisphere Publishing Corporation / Taylor & Francis Group, 1989.

[701] A. J. J. M. Vingerhoets. Biochemical changes in two subjects succumbing to syncope. *Psychosomatic Medicine*, 46(2):95–103, 1984.

[702] M. M. R. Vollenbroek-Hutten, H. J. Hermens, and R. M. H. A. Huis in 't Veld. *Context aware tele-monitoring and tele-treatment services*, volume 114 of *Studies in Health Technology and Informatics*, pages 230–234. Amsterdam, The Netherlands: IOS Press, 2005.

[703] E. Vyzas and R. W. Picard. Affective pattern classification. In *Emotional and Intelligent: The tangled knot of cognition – Papers from the AAAI Fall Symposium*, volume Technical report: FS-98-03, pages 176–182, Orlando, FL, USA, October 23–25 1998. Menlo Park, CA, USA: ŢAssociation for the Advancement of Artificial Intelligence.

[704] J. Wagner, J. Kim, and E. André. From physiological signals to emotions: Implementing and comparing selected methods for feature extraction and classification. In *Proceedings of the IEEE International Conference on Multimedia and Expo (ICME)*, pages 940–943, Amsterdam, The Netherlands, July 6–8 2005.

[705] H. L. Wang and L. Cheong. Affective understanding in film. *IEEE Transactions on Circuits and Systems for Video Technology*, 16(6):689–704, 2006.

[706] L. Wang, T. Gu, X. Tao, H. Chen, and J. Lu. Recognizing multi-user activities using wearable sensors in a smart home. *Pervasive and Mobile Computing*, 7(3):287–298, 2011.

[707] D. Watson and A. Tellegen. Toward a consensual structure of mood. *Psychological Bulletin*, 98(2):219–235, 1985.

[708] T. W. Way, B. Sahiner, L. M. Hadjiiski, and H.-P. Chan. Effect of finite sample size on feature selection and classification: A simulation study. *Medical Physics*, 37(2):907–920, 2010.

[709] A. J. Weigert. *Mixed Emotions: Certain steps toward understanding ambivalence*. SUNY series in the sociology of emotions. Albany, NY, USA: State University of New York Press, 1991.

[710] E. W. Weisstein. *CRC Concise Encyclopedia of Mathematics*. Boca Raton, FL, USA: Chapman & Hall / CRC, 2nd edition, 2003.

[711] J. Weizenbaum. ELIZA – A computer program for the study of natural language communication between man and machine. *Communications of the ACM*, 9(1):36–45, 1966.

[712] C. Wendt and A. P. Petropulu. Pitch determination and speech segmentation using the discrete wavelet transform. In *IEEE International Symposium on Circuits and Systems: Connecting the World*, volume 2, pages 45–48, Atlanta, GA, USA, May 12–15 1996. New York, NY, USA: IEEE Computer Society.

[713] J. Westerink, M. Ouwerkerk, G.-J. de Vries, S. de Waele, J. van den Eerenbeemd, and M. van Boven. Emotion measurement platform for daily life situations. In *Proceedings of the IEEE 3rd International Conference on Affective Computing and Intelligent Interaction, ACII*, volume 1, pages 217–222, Amsterdam, The Netherlands, 10–12 September 2009. IEEE Press.

[714] J. H. D. M. Westerink, M. Krans, and M. Ouwerkerk. *Sensing Emotions: The impact of context on experience measurements*, volume 12 of *Philips Research Book Series*. Dordrecht, The Netherlands: Springer Science+Business Media B.V., 2011.

[715] J. H. D. M. Westerink, B. G. M. M. Majoor, and M. D. Rama. Interacting with infotainment applications: Navigation patterns and mental models. *Behaviour & Information Technology*, 19(2):97–106, 2000.

[716] J. H. D. M. Westerink, M. Ouwerkerk, T. Overbeek, W. F. Pasveer, and B. de Ruyter. *Probing Experience: From assessment of user emotions and behaviour to development of products*, volume 8 of *Philips Research Book Series*. Dordrecht, The Netherlands: Springer Science + Business Media B.V., 2008.

[717] J. Whitehill, G. Littlewort, I. Fasel, M. Bartlett, and J. Movellan. Towards practical smile detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(11):2106–2111, 2009.

[718] S. Whittle, M. Yücel, M. B. H. Yap, and N. B. Allen. Sex differences in the neural correlates of emotion: Evidence from neuroimaging. *Biological Psychology*, 87(3):319–333, 2011.

[719] J. Wickins. The ethics of biometrics: The risk of social exclusion from the widespread use of electronic identification. *Science and Engineering Ethics*, 13(1):45–54, 2007.

[720] J. L. Wilbarger, L. C. Reed, and D. N. McIntosh. Implicit influence of affective postures on the perception of others: You can't show me how i feel. *Emotion*, 11(3):481–491, 2011.

[721] J. F. Wilder. *Stimulus and Response: The Law of Initial Value*. Bristol, UK: John Wright & Sons Ltd., 1967.

[722] B. H. Wixom and P. A. Todd. A theoretical integration of user satisfaction and technology acceptance. *Information Systems Research*, 16(1):85–102, 2005.

[723] J. Wolpe. *Psychotherapy by reciprocal inhibition*. Stanford, CA, USA: Stanford University Press, 1958.

[724] M. Wooldridge. *An introduction to multiagent systems*. Chichester, West Sussex, England: John Wiley & Sons Ltd., 2nd edition, 2009.

[725] S. Wu, T. H. Falk, and W.-Y. Chan. Automatic speech emotion recognition using modulation spectral features. *Speech Communication*, 53(5):768–785, 2011.

[726] G. Wübbeler, M. Stavridis, D. Kreiseler, R.-D. Bousseljot, and C. Elster. Verification of humans using the electrocardiogram. *Pattern Recognition Letters*, 28(10):1172–1175, 2007.

[727] R. Xiao, Q. Zhao, D. Zhang, and P. Shi. Facial expression recognition on multiple manifolds. *Pattern Recognition*, 44(1):107–116, 2011.

[728] S. Xiao, A. Dhamdhere, V. Sivaraman, and A. Burdett. Transmission power control in body area sensor networks for healthcare monitoring. *IEEE Journal on Selected Areas in Communications*, 27(1):37–48, 2009.

[729] R. Xu and D. C. Wunsch. Clustering algorithms in biomedical research: A review. *IEEE Reviews in Biomedical Engineering*, 3(1):1–35, 2010.

[730] R. Xu and D. Wunsch II. Survey of clustering algorithms. *IEEE Transactions on Neural Networks*, 16(3):645–678, 2005.

[731] Y. Xu, D. Zhang, and J.-Y. Yang. A feature extraction method for use with bimodal biometrics. *Pattern Recognition*, 43(3):1106–1115, 2010.

[732] Y. Yacoob and L. S. Davis. Recognizing human facial expressions from long image sequences using optical flow. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(6):636–642, 2006.

[733] Y.-H. Yang, Y.-C. Lin, Y.-F. Su, and H. H. Chen. Music emotion classification: A regression approach. In *Proceedings of the IEEE International Conference on Multimedia and Expo*, pages 208–211, Beijing, China, July 2–5 2007. Piscataway, NJ, USA: IEEE.

[734] G. N. Yannakakis and J. Hallam. Entertainment modeling through physiology in physical play. *International Journal of Human-Computer Studies*, 66(10):741–755, 2008.

[735] S. Yildirim, S. Narayanan, and A. Potamianos. Detecting emotional state of a child in a conversational computer game. *Computer Speech and Language*, 29(1):29–44, 2011.

[736] S. K. Yoo, C. K. Lee, J. Y. Park, N. H. Kim, B. C. Lee, and K. S. Jeong. Neural network based emotion estimation using heart rate variability and skin resistance. *Lecture Notes in Computer Science (Advances in Natural Computation)*, 3610:818–824, 2005.

[737] M. A. Young. Supplementing tests of statistical significance: Variation accounted for. *Journal of Speech and Hearing Research*, 36(4):644–656, 1993.

[738] T. Y. Young and W. H. Huggins. On the representation of electrocardiograms. *IRE Transactions on Bio-Medical Electronics*, 10(3):86–95, 1963.

[739] Z. Zeng, M. Pantic, G. I. Roisman, and T. S. Huang. A survey of affect recognition methods: Audio, visual, and spontaneous expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(1):39–58, 2009.

[740] J. Zhai and A. Barreto. Stress detection in computer users through noninvasive monitoring of physiological signals. *Biomedical Science Instrumentation*, 42:495–500, 2006.

[741] S. Zhang, Q. Huang, S. Jiang, W. Gao, and Q. Tian. Affective visualization and retrieval for music video. *IEEE Transactions on MultiMedia*, 12(6):510–522, 2010.

[742] X. Zhang. A multiscale progressive model on virtual navigation. *International Journal of Human-Computer Studies*, 66(4):243–256, 2008.

[743] Y. Zhang and Q. Ji. Active and dynamic information fusion for facial expression understanding from image sequences. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(5):699–714, 2005.

[744] N. Zheng, Z. Wu, M. Lin, and L. T. Yang. Enhancing battery efficiency for pervasive health-monitoring systems based on electronic textiles. *IEEE Transactions on Information Technology in Biomedicine*, 14(2):350–359, 2010.

[745] J. Z. Zhou and X. H. Wang. Multimodal affective user interface using wireless devices for emotion identification. In *Proceedings of the 27th Annual International Conference of the Engineering in Medicine and Biology Society*, pages 7155–7157, Shanghai, China, September 1–4 2005. Piscataway, NJ, USA: IEEE, Inc.

[746] H. Zhu and R. Rohwer. No free lunch for cross validation. *Neural Computation*, 8(7):1421–1426, 1996.

A

Statistical techniques

Abstract

This appendix will provide a concise introduction to all statistical techniques employed in this monograph. Three unsupervised statistical analyses will be introduced, namely: principal component analysis (PCA), analysis of variance (ANOVA), and linear regression models (LRM). Next, three supervised statistical analysis or machine learning techniques will be introduced, namely: k-nearest neighbors (kNN), artificial neural networks (ANN), and support vector machines (SVM). Last, validation of models will be discussed. In particular, leave-one-out cross validation (LOOCV).

This appendix is based on Section 3 of:

Broek, E.L. van den, Lisý, V., Janssen, J.H., Westerink, J.D.H.M., Schut, M.H., and Tuinenbreijer, K. (2010). Affective Man-Machine Interface: Unveiling human emotions through biosignals. In A. Fred, J. Filipe & H. Gamboa (Eds.), *BioMedical Engineering Systems and Technologies (series: Communications in Computer and Information Science, Vol. 52)*, p. 21–47. Berlin/Heidelberg, Germany: Springer-Verlag. *[invited]*

and on Appendix 1 of:

Broek, E.L. van den, Sluis, F. van der, and Dijkstra, T. (2011). Telling the story and re-living the past: How speech analysis can reveal emotions in post-traumatic stress disorder (PTSD) patients. In J.H.D.M. Westerink, M. Krans, and M. Ouwerkerk (Eds.), *Sensing Emotions: The impact of context on experience measurements (Chapter 10)*, p. 153–180. Series: Philips Research Book Series, Vol. 12. Dordrecht, The Netherlands: Springer Science+Business Media B.V. *[invited]*

A.1 Introduction

In this appendix, we will briefly introduce the techniques used in the research described in this monograph. This appendix is not meant to provide a complete overview or taxonomy of statistical analysis and statistical pattern analysis and machine learning. For this we refer to the many handbooks (e.g., [7, 48, 170, 266, 457, 586, 648, 691]) and survey articles (e.g., [68, 76, 141, 225, 226, 243, 295, 306–308, 310, 347, 348, 394, 468, 602, 690, 729–731]) that have been published on (statistics and) the various statistical pattern analysis and machine learning techniques. With this appendix, I simply hope to improve the readability of the monograph for those readers who are not familiar with all of the statistical techniques applied throughout the monograph. A general glossary of terms which can also help in this has been composed by Kohavi and Provost [355].

First, three unsupervised statistical analyses will be introduced, namely: principal component analysis (PCA), analysis of variance (ANOVA), and linear regression models (LRM). Second, three supervised statistical analysis or machine learning techniques will be introduced, namely: k-nearest neighbors (kNN), artificial neural networks (ANN), and support vector machines (SVM). Third and last, validation of models will be discussed. In particular, leave-one-out cross validation (LOOCV).

In this monograph, PCA is employed in Chapter 7 to enabled the selection of a subset of features for the automatic classification of the emotions. Feature selection / reduction is a crucial phase in the pattern recognition processing pipeline, as it can significantly improve its success, its generalizability, and its efficiency; see also Figure 1.2. It is one of the most frequently used feature selection/reduction techniques employed in automatic emotion recognition.

ANOVA, and its variations, is the most widely distributed statistical test. This is no different for research on emotions. In this monograph, ANOVA is mainly used to determine whether or not there is a significant difference between the means of several emotions. However, in Chapter 7 it is also employed for feature selection, as an alternative for PCA.

LRM enable the mapping of one dependent variable (e.g., an emotion) onto a set of independent variables (e.g., speech and/or biosignal parameters). LRM provide an optimal linear model that describes this mapping. It is one of the most widely used statistical modeling techniques.

k-NN is a simple, intuitive, and elegant supervised machine learning algorithm. Although k-NN is simple and many more advanced machine learning strategies have been introduced, it is still frequently used. This is no different for research on the classification of emotions. k-NN's popularity has many reasons, among which the following triplet: i) Its simplicity makes it both easy to understand and to implement. ii) Throughout time k-NN has established itself as a sort of a baseline for machine learning techniques. iii) Although the algorithm is rather simple, it often performs surprisingly well and, in adapted form, even outperforms some far more advanced machine learning strategies (e.g., SVM. [698]).

ANN are inspired by biological neural networks, as they have some common characteristics, which made them an intriguing machine learning technique. Throughout the years various ANN topologies have been introduced. I will refrain from providing a complete taxonomy of these topologies. Instead I will depict an important class and introduce its most important representant: the Multi-Layer Perceptron (MLP). The MLP is among the most widely used ANN. This is no different for automated emotion classification.

SVM are currently very popular. Linear SVM were introduced in 1963 by Vapnik [688], nonlinear SVM were introduced by Boser, Guyon, and Vapnik [60], and, subsequently, extended with "soft margines" (i.e., a modified maximum margin, which allows mislabeled examples) by Cortes and Vapnik [125]. With the two adaptations in the 90s of the previous century, SVM became a flexible and very powerful class of machine learning techniques. For reasons of brevity, I will restrict the introduction of SVM to the two class problem; however, it should be noted that SVM are suitable to handle multiple classes as well (e.g., see [244]).

In principle, supervised machine learning techniques can be developed that correctly classify 100% of the data. However, the does not answer the question of how they would classify a new set of similar but not identical data. In other words, how well does the classifier generalize? The traditional way to tackle this issue was to split the data set into a train and a test set. However, often, little data is available which reduces the generation of a robust model (i.e., the classifier). To optimally exploit the data available, the principle of cross validation can be exploited. Cross validation was introduced by Stone in the 70s of the previous century [628–630], his two papers are essential reading on this topic. This appendix will introduce leave-one-out cross validation, which is the strongest form of cross validation.

A.2 Principal component analysis (PCA)

Through principal component analysis (PCA), the dimensionality of a data set of interrelated variables can be reduced, preserving its variation as much as possible [319]. For example, Langley, Bowers, and Murray [375] applied PCA to describe the respiratory-induced variability of ECG features, P waves, QRS complexes, and T waves. The variables (e.g., ECG's signal characteristics) are transformed to a new set of uncorrelated but ordered variables: the principal components $\alpha \cdot x$. The first principal component represents, as good as possible, the variance of the original variables. Each succeeding component represents the remaining variance, as good as possible. For a brief introduction on PCA, we refer to [551, Chapter 12].

Suppose we have data that can be represented as vectors x, which consists of n variables. Then, the principal components are defined as a linear combination $\alpha \cdot x$ of the variables of x that preserves the maximum of the (remaining) variance, denoted as:

$$\alpha \cdot x = \sum_{i=0}^{n-1} \alpha_i x_i,$$

where $\alpha = (\alpha_0, \alpha_1, \dots, \alpha_{n-1})^T$. The variance covered by each principal component $\alpha \cdot x$ is defined as:

$$var(\alpha \cdot x) = \alpha \cdot C\alpha,$$

where C is the covariance matrix of x.

To find all principal components, we need to find the maximized $var(\alpha \cdot x)$ for them. Hereby, the constraint $\alpha \cdot \alpha = 1$ has to be taken into account. The standard approach to do so is the technique of Lagrange multipliers. We maximize

$$\alpha \cdot C\alpha - \lambda \left(\sum_{i=0}^{n-1} \alpha_i^2 - 1 \right) = \alpha \cdot C\alpha - \lambda(\alpha \cdot \alpha - 1),$$

where λ is a Lagrange multiplier. Subsequently, we can derive that λ is an eigenvalue of C and α is its corresponding eigenvector.

Once we have obtained the vectors α, a transformation can be made that maps all data x to its principal components:

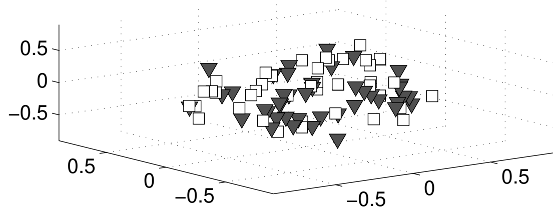
$$x \to (\alpha_0 \cdot x, \alpha_1 \cdot x, \dots, \alpha_{n-1} \cdot x)$$

Note that the principal components are sensitive to scaling. In order to tackle this problem, the components can be derived from the correlation matrix instead of the covariance matrix. This is equivalent to extracting the principal components in the described way after normalization of the original data set to unit variance.

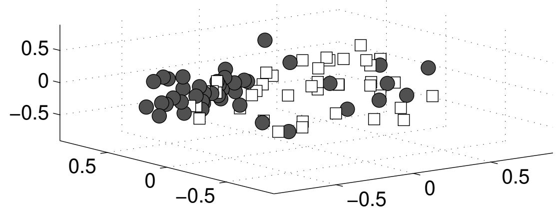
PCA is also often applied for data inspection through visualization, where the principal components are chosen along the figure's axes. Figure A.1 presents such a visualization: for each set of two emotion classes, of the total of four, a plot denoting the first three principal components is presented.

A.3 Analysis of variance (ANOVA)

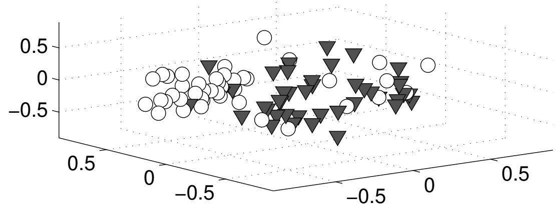
Analysis of variance (ANOVA) is a statistical test to determine whether or not there is a significant difference between the means of several data sets. ANOVA examines the variance



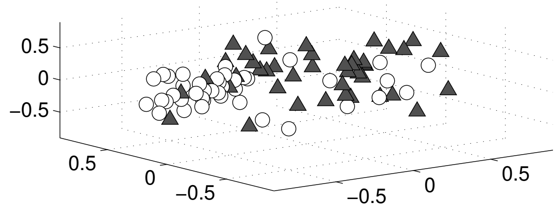
(a) negative (▼) versus mixed (□) emotions



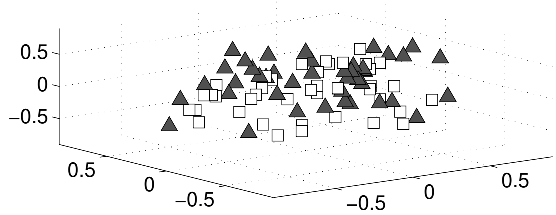
(b) neutral (●) versus mixed (□) emotions



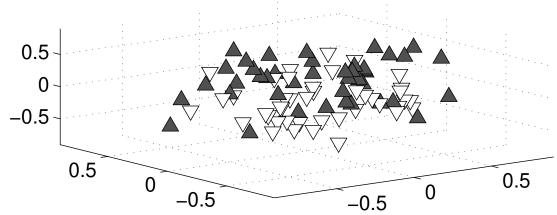
(c) neutral (○) versus negative (▼) emotions



(d) neutral (○) versus positive (▲) emotions



(e) positive (▲) versus mixed (□) emotions



(f) positive (▲) versus negative (▽) emotions

Figure A.1: Visualization of the first three principal components of all six possible combinations of two emotion classes. The emotion classes are plotted per two to facilitate the visual inspection. The plots illustrate how difficult it is to separate even two emotion classes, where separating four emotion classes is the aim. However, note that the emotion category neutral can be best separated from the other three categories: mixed, negative, and positive emotions, as is illustrated in b), c), and d).

of data set means compared to within class variance of the data sets themselves. As such, ANOVA can be considered as an extension of the t-test, which can only be applied on one or two data sets. We will sketch the main idea here. For a more detailed explanation, we refer to [551, Chapter 6].

ANOVA assumes that the data sets are independent and randomly chosen from a normal distribution. Moreover, it assumes that all data sets are equally distributed. These assumptions usually hold with empirical data. Moreover, the test is fairly robust against limited violations.

Assume we have D data sets. For each data set d, the sum t_d and mean $\bar{s}_d$ of all samples are defined as:

$$t_d = \sum_{i=0}^{S-1} x_{id} \quad \text{and} \quad \bar{s_d} = \frac{t_d}{s_d}$$

where x_{id} denotes one data sample and s_d denotes the number of samples of data set d. Subsequently, the grand sum T and the total number of data samples S can be defined as:

$$T = \sum_{d=0}^{D-1} t_d \quad \text{and} \quad S = \sum_{d=0}^{D-1} s_d.$$

The total sum of squares SS (i.e., the quadratic deviation from the mean) can be written as the sum of two independent components:

$$SS_H = \sum_{d=0}^{D-1} \frac{t_d^2}{s_d^2} - \frac{T^2}{S} \quad \text{and} \quad SS_E = \sum_{d=0}^{D-1} \sum_{i=0}^{S-1} x_{id}^2 - \sum_{d=0}^{D-1} \frac{t_d^2}{s_d^2},$$

where indices $_H$ and $_E$ denote hypothesis and error, as is tradition in social sciences. Together with S and D, these components define the ANOVA statistic:

$$F(D-1, S-D) = \frac{S-D}{D-1} \cdot \frac{SS_H}{SS_E},$$

where $D-1$ and $S-D$ can be considered as the degrees of freedom.

The hypothesis that all data sets were drawn from the same distribution is violated if

$$F_\alpha(D-1, S-D) < F(D-1, S-D),$$

where F_α denotes the ANOVA statistic that accompanies chance level α, considered to be acceptable. Often α is chosen as either 0.05, 0.01, or 0.001. Note that as such, ANOVA can also be considered as an inverse k-means clustering. It assumes k clusters and tests the validity of this assumption. It determines whether or not the clusters are independent through an evaluation of between-group variability against within-group variability.

Some conventions in reporting ANOVAs should be mentioned briefly as well. Tests are reported with their degrees of freedom, exact power, and exact level of significance (i.e., α). If α is close to zero, this will be denoted with p $<$ .001, instead of providing an exact α statistic. As measure of effect size (partial) Eta squared (η^2) is often reported, which indicates the proportion of variance accounted for (i.e., a generalization of r/r^2 and R/R^2 in correlation/regression analysis; see also the next subsection) [211, 737]. The threshold for reporting results is $\alpha \leq .050$, results with $.050 < \alpha \leq .100$ are often denoted as trends.

A.4 Linear regression models

A linear regression model (LRM) is an optimal linear model of the relationship between one dependent variable (e.g., the SUD) and several independent variables (e.g., the speech features). A linear regression model typically takes the following form:

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p + \varepsilon, \tag{A.1}$$

where ε represents unobserved random noise, and p represents the number of predictors (i.e., independent variables x and regression coefficients β). The linear regression equation is the result of a linear regression analysis, which aims to solve the following n equations in an optimal fashion:

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} x_{11} & \cdots & x_{1p} \\ x_{21} & \cdots & x_{2p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{np} \end{pmatrix} \times \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}. \tag{A.2}$$

Here, there are n equations for n data points of y. As there are normally more than one solutions to the problem, a least squares method is used to give the optimal solution. Please consult a handbook (e.g., [262]) for more information on the least squares method and its alternatives. A discussion of this topic falls beyond the scope of this appendix.

The following characteristics are used to describe an LRM:

1. Intercept: the value of β_0.
2. Beta (B) and Standard Error (SE): the regression coefficients and standard error of its estimates.
3. Standardized B (β): the standardized Betas, in units of standard deviation of its estimates.
4. T-test (t): a t-test for the impact of the predictor.
5. F-test (F): an ANOVA testing the goodness of fit of the model for predicting the dependent variable.
6. R-square (R^2): the amount of explained variance by the model relative to the total variance in the dependent variable.
7. Adjusted R-square ($\overline{R}^2$): R-square (R^2) penalized for the number of predictors used.

Last, it should be noted that regression models can also fit nonlinear functions. For example, exponential and logistic models with additive error terms can be employed. An in depth discussion of this falls well beyond the scope of this Appendix. For more information on nonlinear regression models, we refer to two classic handbooks [34, 596] as well as a recent handbook [261].

A.5 k-nearest neighbors (k-NN)

k-nearest neighbors (k-NN) is a very intuitive, simple, and frequently applied machine learning algorithm (cf. [4]). It requires only a set of labeled examples (i.e., data vectors), which form the training set [129]. k-NN can fit linear as well as nonlinear functions [698].

Now, let us assume that we have a training set x^l and a set of class labels C. Then, each new vector x_i from the data set is classified as follows:

1. Identify k vectors from x^l that are closest to vector x_i, according to a metric of choice; for example, city block, Euclidean, or Mahalanobis distance. For more information on the choice of k and its importance, I refer to [253].
2. Class c_i that should be assigned to vector x_i is determined by:

   $$c_i = \underset{c \in C}{\operatorname{argmax}} \sum_{i=0}^{k-1} w_i \gamma(c, c_i^l),$$

   where $\gamma(.)$ denotes a Boolean function that returns 1 when $c = c_i^l$ and 0 otherwise and

   $$w_i = \begin{cases} 1 & \text{if } \delta(x_i, x_i^l) = 0; \\ \frac{1}{d(x_i, x_i^l)^2} & \text{if } \delta(x_i, x_i^l) \neq 0, \end{cases}$$

   where $\delta(.)$ denotes the distance between vectors x_i and x_i^l. Note that, if preferred, the factor weight can be simply eliminated by putting $w_i = 1$.
3. If there is a tie of two or more classes $c \in C$, vector x_i is randomly assigned to one of these classes.

The algorithm presented applies to k-NN for weighted, discrete classifications, as was applied in the current research. However, a simple adaptation can be made to the algorithm, which enables continuous classifications.

Also, please note that (distance-weighted) k-NN can be generalized to estimate regression models [157], which is often denoted as (k-)NN regression [465]. For more information on these and other issues, we refer to the various freely available tutorials and introductions that have been written on k-NN.

A.6 Artificial neural networks (ANN)

The human brain has about 10^{11} neurons and each neuron may have 10^4 synaptic inputs and input weights. In contrast, artificial neural networks (ANN) consist of only a few dozen units (or neurons or perceptrons) and their number of inputs is even less [266]. Many ANN learn by adapting the synaptic weight values against each other when training examples are presented. However, where in the brain would the computing process reside, which would execute synaptic weight adjusting algorithms and where would these algorithms have come from? The evolutionary feasibility of these kinds of algorithms can be seriously questioned. Nevertheless, ANN are frequently claimed to have similar behavior to biological neural networks. However, it should be noted that this is a claim that can hardly be justified. Nevertheless, ANN have proved their use for a range of pattern recognition and machine learning applications. One of their appealing characteristics is that they can fit nonlinear functions. Moreover, ANN have a solid theoretical basis [48, 457].

ANN consist of a layer of input units, one or more layers of hidden units, and a layer of output units. These units are connected with a weight w_{ij}, which determines the transfer of unit u_i to unit u_j. The activation level of a unit u_j is defined as:

$$a_j(t+1) = f(a_j(t), i_j(t)),$$

where t denotes time, $f(.)$ is the activation function that determines the new activation based on the current state $a(t)$ and its effective input, defined as:

$$i_j(t) = \sum_{i=0}^{U_j-1} a_i(t) w_{ij}(t) + \tau_j(t),$$

where $\tau_j(t)$ is a certain bias or offset and U_j denotes the number of units from which a unit u_j can receive input. Note that at the input layer of an ANN, the input comes from the environment; then, i is the environment instead of another unit.

On its own, each neuron (or perceptron) of an ANN can only perform a simple task. In contrast, a proper network of units can approximate any function [48, 170, 266, 395]. Moreover, ANN cannot only process input, they can also learn from their input, either supervised or unsupervised. Although various learning rules have been introduced for ANN, most can be considered as being derived from Hebb's classic learning rule:

$$\Delta w_{ij} = \eta a_i a_j,$$

which defines the modification of the weight of connection (u_i, u_j). η is a positive constant. Its rationale is that w_{ij} should be increased with the simultaneous activation of both units and the other way around.

Various ANN topologies have been introduced. The most important ones are recurrent and feed-forward networks, whose units respectively do and do not form a directed cycle through feedback connections. In the current research, a feed-forward networks have been applied: the classic Multi-Layer Perceptron (MLP), as is more often used for emotion recognition purposes; see also Table 2.4. It incorporated the often adopted sigmoid-shaped function applied to $f(.)$:

$$\frac{1}{1+e^{-a_j}}$$

Throughout the 60 years of their existence, a broad plethora of ANN have been presented, varying on a range of aspects; for example, their topology, learning rules, and the choice of either synchronous or asynchronously updating of its units. More information on ANN can be found in various introductions to ANN.

A.7 Support vector machine (SVM)

Using a suitable kernel function, a support vector machine (SVM) ensures the division of a set of data into two classes, with respect to the shape of the classifier and the incorrect classification of the training samples [96, 125]. For a recent concise survey on SVM as well a review on optimization techniques for SVM, I refer to [602]. The main idea of SVM can best be explained with the example of a binary linear classifier.

Let us define our data set as:

$$D = \left\{(x_i, c_i) | x_i \in \mathbb{R}^d, c_i \in \{-1, 1\}\right\} \text{ for } i = 0, 1, \ldots, N-1,$$

where x_i is a vector with dimensionality d from the data set, which has size N. c_i is the class to which x_i belongs. To separate two classes, we need to formulate a separating hyperplane $w \cdot x = b$, where w is a normal vector of length 1, x is a feature vector, and b is a constant.

In practice, it is often not possible to find such a linear classifier. In this case, the problem can be generalized. Then, we need to find w and b so that we can optimize

$$c_i(w \cdot x_i + b) \leq \xi_i,$$

where ξ_i represents the deviation (or error) from the linearly separable case.

To determine an optimal plane, the sum of ξ_i must be minimized. The minimization of this parameter can be solved by Lagrange multipliers α_i. From the derivation of this method, it is possible to see that often most of the α_is are equal to 0. The remaining relevant subset of the training data x is denoted as the support vectors. Subsequently, the classification is performed as:

$$f(x) = \mathrm{sgn}\left(\sum_{i=0}^{S-1} c_i \alpha_i x \cdot x_i + b\right),$$

where S denotes the number of support vectors.

For a non-linear classification problem, we can replace the dot product by a non-linear kernel function. This enables the interpretation of algorithms geometrically in feature spaces non-linearly related to the input space and combines statistics and geometry. A kernel can be viewed as a (non-linear) similarity measure and can induce representations of the data in a linear space. Moreover, the kernel implicitly determines the function class, which is used for learning [586].

The SVM introduced here classified samples into two classes. In the case of multiple classes, two approaches are common: 1) for each class, a classifier can be built that separates that class from the other data and 2) for each pair of classes, classifiers can be built. With both cases, voting paradigms are used to assign the data samples x_i to classes c_i. For more information on SVM, [48, 586] can be consulted.

A.8 Leave-one-out cross validation (LOOCV)

Assume we have a classifier that is trained, using a part of the available data set: the training data. The training process optimizes the parameters of a classifier to make it fit the training data. To validate the classifier's performance, an independent sample of the same data set has to be used [48, 457]. Cross validation deviates from the general validation scheme since it enables the validation of a classifier without the need of an explicit validation set. As such, it optimizes the size of the data set that can be used as training data.

Various methods of cross validation have been introduced. In this section, we will introduce leave-one-out cross validation (LOOCV), a method frequently used to determine the performance of classifiers. LOOCV is typically useful and, consequently, used in the analysis of (very) small data sets. It has been shown that LOOCV provides an almost unbiased estimate of the true generalization ability of a classifier. As such, it provides a good model selection criterion.

Assume we have a classifier (e.g., k-NN, a SVM, or an ANN) of which we want to verify the performance on a particular data set. This data set contains (partly) data samples x_i with known correct classifications c_i^l. Then, the classifier's performance can be determined through LOOCV, as follows:

1. $\forall_i$ train a classifier C_i with the complete data set x, except x_i.

2. $\forall_i$ classify data sample x_i to a class c_i, using classifier C_i.
3. Compute the average error of the classifier through

$$\mathcal{E} = \frac{1}{D} \operatorname*{argmax}_{c \in C} \sum_{i=0}^{D-1} \gamma(c_i, c_i^l),$$

where D denotes the number of data samples and $\gamma(.)$ denotes a Boolean function, which returns 1 if $c_i = c_i^l$ and 0 otherwise. Note that $\frac{1}{D}$ can be omitted from the formula if no comparisons are made between data sets (with different sizes).

Instead of one data sample x_i, this validation scheme also allows a subset of the data to be put aside. Such a subset can, for example, consist of all data gathered from one person. This enables an accurate estimation of the classification error $\mathcal{E}$ on this unknown person.

The processing scheme as presented here can be adapted in various ways. For example, in addition to the Boolean function $\gamma(.)$, a weight function could be used that expresses the resemblance between classes. Hence, not all misclassifications would be judged similarly.

All results reported in Chapter 7 were determined through LOOCV, if not otherwise specified. For more information on cross validation, we refer to [177, Chapter 17] and as more general reference works to [48, 516]. For an interesting discussion on the topic we refer to a series of articles [231, 555, 746].

Although only rarely applied, the principle of cross validation (e.g., LOOCV) can also be applied on unsupervised learning techniques. For example, Picard and Cook applied cross validation on LRM [519]. Similarly, cross validation could be employed on PCA and ANOVA, although its implications will be a little different.

Summary

Slowly computers are becoming dressed, huggable and tangible operating interfaces. They are being personalized and are expected to understand more of their users' feelings, emotions, and moods. Consequently, concepts, such as emotions, which were originally the playing field of philosophers, sociologists, and psychologists, have also become entangled with computer science. In 1997, Picard baptized the topic *affective computing*. Moreover, she identified biosignals as an important covert channel to capture emotions, in addition to channels such as speech and computer vision. This monograph explores several factors that have been posed to be of key importance to *affective computing*, in particular to *affective signal processing (ASP)*: the use of biosignals for emotion-aware systems. It is divided into five parts: i) a *prologue*, ii) basic research on *baseline-free ASP*, iii) basic research on *bi-modal ASP*, iv) three studies *towards affective computing*, and v) an *epilogue*. Additionally, an appendix provides a description of all statistical and pattern recognition techniques used. Here is a concise summary of each of the five parts.

In the introduction, Chapter 1 of the prologue (*Part I*), a brief introduction of emotion theory, the field of *affective computing*, *A*SP, and their relevance for computer science is provided. Human-Computer Interaction, Artificial Intelligence, and health informatics are described. The monograph's working model, a closed loop model (i.e., a control system with an active feedback loop), is introduced and its signal processing and classification components are described. A concise overview of the biosignals investigated is given. In Chapter 2 a review of *a*ffective computing is presented, with an emphasis on *ASP* using biosignals. The conclusion of this chapter is that *ASP* lacks the progress it needs. Possible angles of view that can aid *ASP*'s progress are explored in the next three parts.

In *Part II* two basic studies on baseline-free *ASP* using statistical moments are presented. These two studies address a number of key issues for *ASP*. Chapter 3 covers research for which dynamic real world stimuli (i.e., movie scenes) were used to induce emotions. The ElectroMyoGraphy (EMG) of three facial muscles was recorded, which is often done to establish a ground truth measurement. In addition, the participants' ElectroDermal Activity (EDA) was recorded. EDA is a robust well documented biosignal that reveals the level of experienced arousal. In Chapter 4 analyses on the same data set as in Chapter 3 are reported. The studies differ in the choice of time windows, which enabled research towards the impact and usage of this parameter for *ASP*. Moreover, events in the movie scenes were directly linked to affective responses.

Part III Two studies are presented that employed bi-modal *ASP* by the rare combination of ElectroCardioGram (ECG) and speech. These studies only differ from each other with respect to the stimuli that were used for emotion elicitation, which has recently been shown to be a factor of importance [8]. The research presented in these two chapters also assessed the influence of emotion representations by analyzing the obtained data using both the dimensional valence-arousal model and the six basic emotions. Moreover, the impact of the environment (or context), the personality traits neuroticism and extroversion, and demographics on *ASP* was explored. In Chapter 5 research is reported that employed a (or perhaps even *the*) reference set for *affective computing*: Lang, Bradley, and Cuthbert's (1994) International Affective Picture System (IAPS). In Chapter 6 a study is presented that used the same set of stimuli (i.e., movie fragments) as was used in the research described in Chapters 3 and 4. This enabled a comparison of static versus dynamic stimuli and, as such, assessed their validity.

Part IV consists of three chapters that present studies that work *towards affective computing*. First, in the research in Chapter 7, a complete signal processing + classification processing pipeline for *ASP* is executed on the data already presented in Chapters 3 and 4. Several preprocessing strategies and automatic classifiers are explored. Second, in Chapter 8, two clinical case studies on *ASP* are presented that aim to explore the feasibility of Computer Aided Diagnosis (CAD) for patients suffering from a post-traumatic stress disorder (PTSD). Third, in Chapter 9, the data of the studies presented in Chapter 8 are used to develop a complete signal processing + pattern recognition processing pipeline, similar to the one presented in Chapter 7. As such, this chapter explores the feasibility of the envisioned emotion-aware systems, in this case: *A*SP-based Computer-Aided Diagnosis (CAD) for mental health care.

This monograph's epilogue, *Part V*, consists of two chapters. In the first one, Chapter 10, the lessons learned from the research presented in the previous chapters is described. It formulates a set of prerequisites and guidelines of which the author hopes that it can serve as a user manual for other researchers who are interested in research on *ASP*. In this manual, the following issues are discussed: physical sensing characteristics, temporal construction, normalization, context, validation, triangulation, and user identification. In the second and last chapter of this part, Chapter 11, the monograph ends with a wrap-up of the work, which is followed by a historical reflection. Next, a triplet of applications is presented that is (almost) ready to be brought to the market here and now, which are followed by two possible future applications. This monograph closes with a brief conclusion: The work presented in this monograph revealed several factors of importance for *ASP*, which helps the scientific community to understand *ASP* better. Moreover, I expect that the manual that resulted from the work presented in this monograph will guide future research on *ASP* to higher levels.

Samenvatting

Langzaam maar zeker veranderen computers, ze krijgen een andere vorm dan de klassieke PC kast of een laptop, ze reageren op aanrakingen, kunnen worden aangekleed, en zijn aaibaar. Ze worden persoonlijker en we verwachten dat ze onze gevoelens, emoties en stemmingen begrijpen. Zo kon het ook gebeuren dat een fenomeen als emoties, oorspronkelijk vooral voer voor filosofen, sociologen en psychologen, werd omarmd door informatici. In 1997 noemde Picard de automatische herkenning van emoties door computers *affective computing*. In aanvulling op spraak en visuale perceptie, identificeerde zij physiologische signalen als een belangrijke impliciet kanaal om emoties te identificeren. Dit boek verkent een aantal factoren waarvan gesteld wordt dat deze belangrijk zijn voor *affective computing*, in het bijzonder voor *affective signal processing (ASP)*: het gebruik van fysiologische signalen voor empatische systemen. Het boek is opgedeeld in vijf delen: i) een *proloog*, ii) onderzoek naar *ongecorrigeerde ASP*, iii) onderzoek naar *bi-modale ASP*, iv) drie studies *naar empatische systemen*, en v) een *epiloog*.

In de introductie, Hoofdstuk 1 van de proloog (*Deel I*), wordt een korte introductie gegeven van emotie theorie, het veld *affective computing*, *A*SP en hun relevantie voor informatica. Mens-machine interactie, kunstmatige intelligentie en medische informatica worden beschreven. Het werkmodel van dit boek, een closed loop model (een systeem met een actieve feedback lus), wordt geïntroduceerd en haar signaalverwerkings- and classificatiecomponenten beschreven. Tevens wordt een kort overzicht gegeven van de fysiologische signalen die zijn onderzocht. In hoofdstuk 2 wordt een overzicht gepresenteerd van het reeds bestaande onderzoek naar *a*ffective computing, met nadruk op *ASP* dat gebruik maakt van fysiologische signalen. De conclusie van dit hoofdstuk is dat *ASP* de vooruitgang ontbeert die het nodig heeft. Mogelijke insteken die de vooruitgang in *ASP* kunnen helpen worden in de volgende drie delen verkend.

In *Deel II* worden twee onderzoeken gepresenteerd naar ongecorrigeerde *ASP* door middel van statistische momenten. Hierin werd een aantal belangrijke aspecten van *ASP* onderzocht. In hoofdstuk 3 wordt onderzoek beschreven waarin gebruik werd gemaakt van dynamische stimuli (i.e., filmfragmenten) om emoties los te maken. De ElectroMyoGraphy (EMG) van drie gezichtsspieren van de deelnemers werd opgenomen; dit wordt vaak gedaan om een zgn. *ground truth* te bepalen. Ook werd de ElectroDermal Activity (EDA) van de deelnemers opgenomen. EDA is een robuust, goed gedocumenteerd fysiologisch signaal dat registreert in hoeverre men zich opwindt. In Hoofdstuk 4 worden analyses beschreven van dezelfde data set als in Hoofdstuk 3. Deze onderzoeken verschilden alleen in de keuze van de tijdvakken, wat het mogelijk maakte de impact en het gebruik van deze parameter voor *ASP* te bepalen. De korte tijdvakken zoals gebruikt in Hoofdstuk 4 maakte het ook mogelijk de gebeurtenissen in de filmfragmenten te koppelen aan affectieve reacties.

In *Deel III* worden twee studies gepresenteerd die bi-modale *ASP* toepassen, namelijk de zeldzame combinatie van het ElectroCardioGram (ECG) en het spraaksignaal. Deze studies verschilden enkel voor wat betreft de stimuli die werden gebruikt om emoties los te maken, hetgeen zeer recent een belangrijke factor bleek te zijn [8]. In het onderzoek werd tevens gekeken naar de invloed van emotiemodellen op het analyseren van de verkregen data, waarbij zowel gebruik werd gemaakt van het dimensionale *valence-arousal model* als van de zes basisemoties. Daarnaast werd ook de invloed van de omgeving (of context), de persoonlijkheidskarakteristieken neuroticisme en extroversie, en demografische factoren op *ASP* onderzocht. In Hoofdstuk 5 wordt onderzoek beschreven dat gebruik maakte van één (of misschien wel *de*) referentieset voor *affective computing*: het *International Affective Picture System (IAPS)* van Lang, Bradley, en Cuthbert (1994). In Hoofdstuk 6 wordt een onderzoek behandeld dat gebruik maakte van dezelfde set stimuli (i.e., film fragmenten) als die gebruikt zijn in het onderzoek beschreven in de Hoofdstukken 3 en 4. Dit maakte een vergelijking tussen statische en dynamische stimuli mogelijk en zo kon ook de validiteit van beide sets worden bepaald.

Deel IV bestaat uit drie hoofdstukken waarin onderzoek *naar empatische systemen* wordt beschreven. In Hoofdstuk 7 wordt beschreven hoe de complete signaalverwerkings + classificatie lijn voor *ASP* is uitgevoerd op de data die reeds gepresenteerd was in Hoofdstukken 3 en 4. Verscheidene voorbewerkingen en classificatie-algoritmes werden toegepast. In Hoofdstuk 8 worden twee klinische onderzoeken gepresenteerd die als doel hadden te bezien of een door de computer ondersteunde diagnose voor patiënten met een posttraumatisch stresssyndroom mogelijk is. In Hoofdstuk 9 wordt de data van de onderzoeken uit Hoofdstuk 8 gebruikt om de complete signaalverwerkings + classificatie lijn voor *ASP*, zoals geïntroduceerd in Hoofdstuk 7, toe te passen. Als zodanig verkent dit hoofdstuk de haalbaarheid van empatische systemen, in dit geval op *A*SP-gebaseerde Computer-Aided Diagnosis (CAD) voor de geestelijke gezondheidszorg.

Het epiloog van dit boek, *Deel V*, bestaat uit twee hoofdstukken. In het eerste hoofdstuk, Hoofdstuk 10, komen de lessen aan bod die geleerd zijn van het onderzoek uit de voorgaande hoofdstukken. Een set voorwaarden en richtlijnen is geformuleerd die kunnen dienen als handleiding voor collega's die, net als de auteur, geinteresseerd zijn in onderzoek naar *ASP*. In deze handleiding komen de volgende onderwerpen aan bod: fysische karakteristieken, temporele aspecten, normalisatie, context, validatie, triangulatie en gebruikersidentificatie. In het tweede en laatste hoofdstuk van dit deel, Hoofdstuk 11, wordt het werk uit dit boek samengevat en wordt hierop gereflecteerd vanuit een historisch oogpunt. Vervolgens wordt eerst een drietal toepassingen van *ASP* gepresenteerd die (bijna) klaar zijn voor de markt en vervolgens worden twee mogelijke toekomstige toepassingen beschreven. Als afsluiting van het boek wordt geconcludeerd dat het onderzoek zoals beschreven in dit boek verschillende belangrijke aspecten van *ASP* heeft geïdentificeerd, hetgeen de wetenschappelijke gemeenschap kan helpen *ASP* beter te begrijpen. Verder verwacht ik dat de onderzoekshandleiding die resulteerde uit mijn onderzoek toekomstig onderzoek naar *ASP* naar een hoger niveau kan brengen.

Dankwoord

Wenen, Oostenrijk, 1 augustus 2011

Tijdens mijn afstuderen, rond de eeuwwisseling, onderzocht ik in hoeverre de ernst van psychische trauma's werden gereflecteerd in het spraaksignaal. Dit onderzoek verdiende een passend vervolg maar ik besloot een andere richting te verkennen: *cognitive computer vision*, waarin ik promoveerde. Mijn afstudeeronderwerp heeft mij in die tijd echter nooit losgelaten. Maar meer en meer en keer op keer kwam ik er telkense weer achter dat het zo veel complexer was dan ik al dacht. Met dit boek heb ik geprobeerd enige structuur aan te brengen in het complexe domein van emotieherkenning door computers. Dit boek mag als niet meer dan een aanzet worden beschouwd, echter ook dit bleek al een enorme uitdaging te zijn. Een uitdaging waarbij ik steun heb gehad van veel mensen en ook continu heb mogen samenwerken met tal van mensen.

Ik heb drie begeleiders gehad die mij ieder op hun eigen manier fantastisch begeleid hebben. Ton, bij jou is het allemaal begonnen, meer dan 10 jaar geleden. Alsnog is er dan dit boek, misschien veel te laat maar beter laat dan nooit. Ik zal je nooit genoeg kunnen bedanken voor je onvoorwaardelijke begeleiding en ook steun door al die jaren heen. Joyce, zonder het toen te weten, betekende onze ontmoeting 5 à 6 jaar geleden de definitieve doorstart van mijn onderzoekslijn, waarvan het product hier nu ligt. In al die jaren hebben wij buitengewoon prettig samengewerkt. Ook enorm bedankt voor je snelle reacties, je nauwkeurigheid, je doortastendheid, al het werk dat je hebt verzet en ook voor je oprechtheid. Anton, ik liep bij jou binnen met het gekke idee om een tweede keer te gaan promoveren. Je hebt mij met open armen ontvangen, me je volledige vertrouwen gegeven en later een plaats in je groep gegeven. De vanzelfsprekendheid waarmee je dit hebt gedaan vind ik tot op de dag van vandaag heel bijzonder.

Dear committee members, I conceive it as an honor that you have been willing to judge this monograph. Geachte prof. Apers fantastisch dat u tijd heeft kunnen vrijmaken om als commissielid zitting te nemen. Prof. Esposito, dear Anna, thank you so much for your dedication, for you detailed comments! They have improved this monograph substantially and have been a source of inspiration for me, while finishing this monograph. Geachte prof. Hermens, als expert in twee aan dit boek aansluitende gebieden, beschouw ik het als een eer dat dit boek aan uw criteria voor een promotie heeft voldaan. Prof. Hoenkamp, beste Eduard, fantastisch om je erbij te hebben! Heel erg bedankt voor je immer positief kritische en multidisciplinaire blik op mijn werk. Prof. Schomaker, beste Lambert, net als Eduard, was jij aanwezig bij het ontstaan van dit werk. Wat ik toen niet wist, maar waar ik door de jaren heen achter kwam, is dat jij > 25 jaar geleden al prachtig onderzoek hebt gedaan op het gebied van dit proefschrift. Ik beschouw het als een eer dat dit werk aan jouw standaarden voor een promotie heeft voldaan.

Door de jaren heen heb ik met veel mensen mogen samenwerken. Zonder hen lag dit boek er nu zeker niet. Willem, al vond je mijn onderzoek eigenlijk te divers, je hebt mij wel de ruimte gegeven om mijn eigen onderzoek te doen; heel veel dank hiervoor. Marleen, het was op de zomerfeesten in (ik denk) 2005 dat jouw afstuderen ter sprake kwam. Dit gesprek markeerde de herstart van jouw afstuderen en de doorstart van mijn onderzoek. Vanaf dat moment ging het allemaal heel snel en kwam alles meer dan op z'n pootjes terecht. :) Bedankt voor de bijzonder prettige en efficiente samenwerking. Marjolein, via Marleen, kwam je bij mij (en daarmee bij Joyce) uit. Ook jij bedankt voor de prettige samenwerking. Joris, via Marjolein, kwam jij vervolgens weer bij mij terecht. De afgelopen jaren hebben hebben we bijzonder intensief, efficient maar ook prettig samengewerkt. Ik hoop dat we dit nog even kunnen volhouden! Viliam thank you for your work conducted 5 years ago and thank you for the pleasant cooperation recently. Frans, bedankt voor je leergierigheid, de prettige samenwerking en voor het niet te beroerd zijn om even door te douwen, als het nodig was. Ik kan alleen maar zeggen dat ik hoop dat we nog lang op deze manier door kunnen gaan. :) Jennifer, thank you for the pleasant cooperation and for your commitment to the prerequisites articles. Thank you also for sharing anecdotes, memories, and knowledge, for which I'm all very grateful. Lynn, heel, heel erg véél dank voor je talloze correcties en de fantastische service hierbij. Maar vooral ook voor het geduld hierbij, de snelheid waarmee je mijn stukken corrigeerde en je (dappere) pogingen om me nog wat te leren. Dames van het HMI-secretariaat, heel erg bedankt voor alle support en de snelle reacties op al m'n mailtjes.

Vrienden en familie bedankt voor de steun en vooral ook voor de ontspanning. In het bijzonder, Brommelhoofdje bedankt voor de Kaplan & Sadock, Van Wees B.V. bedankt voor de "printing facilities" en Selma en Arjan bedankt voor B&B in A'dam. Ad enorm bedankt voor de prachtige kaft en boekenlegger die je ook dit boek hebt gegeven. Je hebt m'n gedachten prachtig weten te visualiseren. Oma, al leest u dit dan niet meer, toch enorm bedankt voor de mooie discussies over het werk wat in dit boek is beschreven, voor uw vertrouwen en voor de prachtige boeken, in het bijzonder die van Paul Ekman en Frans de Waal. Papa, mama, Anntje en Rick en Ton en Karin bedankt voor jullie onvoorwaardelijke steun! Lieve Winnie bedankt voor al je hulp, je altijd kritische blik en vooral ook voor je onvoorwaardelijke steun, ook op die momenten waarop het eigenlijk weer eens tè druk was. Zonder jouw steun had dit boek er nu niet gelegen. Fantastisch Tijntje van me, jij zorgde er altijd wel voor dat papa aandacht voor je had; en terecht! Het gaf me de broodnodige afleiding door regelmatig met je te "klieren" of samen een Duplo huis te bouwen. :) Dit "hobby project" is nu afgesloten, dat zal thuis allicht wat meer tijd en rust brengen.

Egon

Curriculum Vitae

Egon L. van den Broek was born on August 22, 1974 in Nijmegen, The Netherlands (NL). He obtained a M.Sc. (2001) in Artificial Intelligence and a Ph.D. (2005) on Human-Centered Content-Based Image Retrieval, both at the Radboud University Nijmegen, NL. His main research areas are human-centered computing (or human media interaction) and intelligent systems, with a specific interest in cognitive computer vision and affective computing. However, both his work and interests comprise many more branches of science and engineering (e.g., ambient intelligence, multimedia, and computational geometry).

Throughout the years, he has guided 50+ students on B.Sc., M.Sc., and Ph.D. level as well as students conducting a minor and post-doctoral students. He has been keynote and invited speaker on several conferences, has published 150+ scientific articles, holds 5 patent applications, (co-)developed several systems, and received various awards. Currently, he is assistant professor at both the University of Twente (NL) and the Radboud University Medical Center Nijmegen, NL. In addition, via Human-Centered Computing Consultancy, he has been and is consultant for various companies and institutes (e.g., Philips and the United Nations Office of Drugs and Crime, UNODC).

Egon holds several additional positions. He is external expert for Expertises Agence Nationale de la Recherche (ANR; France), Innovation by Science and Technology (IWT; Belgium), and The British Academy (UK). He is a member of 30+ national and international boards and program committees. Egon is founding editor-in-chief of the Pan Stanford Series in Artificial Intelligence and editorial board member of the encyclopedia of Interaction-Design.org. He is associate editor of the journals Behaviour & Information Technology and Health and Technology. Further, he is a member of the editorial board of the journals: Central European Journal of Computer Science, Journal of Ambient Intelligence and Smart Environments, Journal of Brain, Face and Emotion, and Journal of Usability Studies.

I: http://www.human-centeredcomputing.com/
I: http://www.cognitive-engineering.org/
E: vandenbroek@acm.org

Publications and Patents: A selection

A selection of publications and patent applications directly related to the work presented in this thesis. A complete list of publications is available upon request.

Publications

Journal & Magazine papers

50. **Broek, E.L. van den**, Janssen, J.H., Zwaag, M.D. van der, Healey, J.A., & Westerink, J.H.D.M. Affective Signal Processing (ASP): A user manual. *ACM Computing Surveys [in preparation]*

49. **Broek, E.L. van den**, Sluis, F. van der, & Dijkstra, T. (2011/2). Modeling an acoustic profile for ubiquitous stress assessment. *Personal and Ubiquitous Computing, 15/6. [Special issue: Mental health and the impact of ubiquitous technologies; in press]*

48. Janssen, J.H., **Broek, E.L. van den**, & Westerink, J.D.H.M. (2011/2). Tune in to your emotions: A robust personalized affective music player. *User Modeling and User-Adapted Interaction, 21/2. [in press]*

47. **Broek, E.L. van den** (2011). Ubiquitous emotion-aware computing. *Personal and Ubiquitous Computing, 15. [in press]*

46. Zwaag, M.D. van der, Westerink, J.D.H.M., & **Broek, E.L. van den** (2011). Emotional and psychophysiological responses to tempo, mode, and percussiveness. *Musicae Scientiae, 15(2)*, 250–269.

45. **Broek, E.L. van den** (2011). Safeguards in a world of ambient intelligence: A social, economic, legal, and ethical perspective. *Journal of Ambient Intelligence and Smart Environments, 3(2)*, 179–180.

44. **Broek, E.L. van den** (2011). Multimedia techniques for device and ambient intelligence: A continuing endeavor. *Journal of Ambient Intelligence and Smart Environments, 3(1)*, 67–68.

43. Bialoskorski, L.S.S., Westerink, J.H.D.M., & **Broek, E.L. van den** (2010). Experiencing affective interactive art. *International Journal of Arts and Technology, 3(4)*, 341–356. *[invited]*

42. **Broek, E.L. van den** (2010). Beyond Biometrics. *Procedia Computer Science, 1(1)*, 2505–2513. *[invited]*

41. **Broek, E.L. van den** (2010). Robot nannies: Future or fiction? *Interaction Studies, 11(2)*, 274–282.

40. **Broek, E.L. van den**, Sluis, F. van der, & Schouten, Th.E. (2010). User-centered digital

preservation of multimedia. *European Research Consortium for Informatics and Mathematics (ERCIM) News, No. 80 (January)*, 45-47. *[Special issue: Digital Preservation]*

39. **Broek, E.L. van den** & Westerink, J.H.D.M. (2009). Considerations for emotion-aware consumer products. *Applied Ergonomics, 40(6)*, 1055–1064. *[Special issue: Psychophysiology in Ergonomics]*

38. **Broek, E.L. van den**, Schut, M.H., Westerink, J.H.D.M., & Tuinenbreijer, K. (2009). Unobtrusive Sensing of Emotions (USE). *Journal of Ambient Intelligence and Smart Environments, 1(3)*, 287–299. *[Thematic Issue: Contribution of Artificial Intelligence to AmI]*

37. Drunen, A. van, **Broek, E.L. van den**, Spink, A.J., & Heffelaar, T. (2009). Exploring workload and attention measurements with uLog mouse data. *Behavior Research Methods, 41(3)*, 868–875. *[invited]*

36. **Broek, E.L. van den**, Pieters, J., Bruikman, H., Sluis, F. van der, Drunen, A. van, & Luursema, J.-M. (2009). Cognitieve Ergonomie. *Perspectief, 4(3)*, 16–19. *[invited]*

35. Bialoskorski, L.S.S., Westerink, J.H.D.M., and **Broek, E.L. van den** (2009). Mood Swings: Design and Evaluation of Affective Interactive Art. *New Review of Hypermedia and Multimedia, 15(2)*, 173–191.

34. Westerink, J., **Broek, E.L. van den**, Herk, J. van, Tuinenbreijer, K., and Schut, M. (2006). Statistical processing of facial electromyography (EMG) signals in emotional film scenes. *Psychophysiology, 43 (S1)*, S106.

33. **Broek, E.L. van den** (2004). Emotional Prosody Measurement (EPM): A voice-based evaluation method for psychological therapy effectiveness. *Studies in Health Technology and Informatics (Medical and Care Compunetics), 103*, 118–125.

32. **Broek, E.L. van den** (2003). A stress marker in speech. *Toegepaste Taalwetenschap in Artikelen, 69(1)*, 143–153 & 172.

Books & Book chapters

31. **Broek, E.L. van den** (2012). *Handbook of Affective Computing.* Series: Pan Stanford Series in Artificial Intelligence. Pan Stanford Publishing. *[in preparation]*

30. **Broek, E.L. van den** (2011). Commentary on Affective Computing by Kristina Höök. In M. Soegaard (Ed.), `interaction-design.org Encyclopedia`. *[in preparation]*

29. **Broek, E.L. van den**, Sluis, F. van der, and Dijkstra, T. (2011). Telling the story and re-living the past: How speech analysis can reveal emotions in post-traumatic stress disorder (PTSD) patients. In J.H.D.M. Westerink, M. Krans, and M. Ouwerkerk (Eds.), *Sensing Emotions: The impact of context on experience measurements (Chapter 10)*, p. 153–180. Series: Philips Research Book Series, Vol. 12. Dordrecht, The Netherlands:

Springer Science+Business Media B.V. *[invited]*

28. **Broek, E.L. van den**, Lisý, V., Janssen, J.H., Westerink, J.D.H.M., Schut, M.H., and Tuinenbreijer, K. (2010). Affective Man-Machine Interface: Unveiling human emotions through biosignals. In A. Fred, J. Filipe & H. Gamboa (Eds.), *BioMedical Engineering Systems and Technologies (series: Communications in Computer and Information Science, Vol. 52)*, p. 21–47. Berlin/Heidelberg, Germany: Springer-Verlag. *[invited]*

27. Westerink, J.H.D.M., **Broek, E.L. van den**, Schut, M.H., Herk, J. van, and Tuinenbreijer, K. (2008). Computing emotion awareness through galvanic skin response and facial electromyography. In J.H.D.M. Westerink, M. Ouwerkerk, T. Overbeek, W.F. Pasveer, and B. de Ruyter (Eds.), *Probing Experience: From Academic Research to Commercial Propositions (Part II: Probing in order to feed back), Chapter 14*, p. 137–150. Series: Philips Research Book Series , Vol. 8. Dordrecht, The Netherlands: Springer Science + Business Media B.V. *[invited]*

Full papers in proceedings (peer reviewed)

26. Sluis, F. van der, **Broek, E.L. van den**, & Dijkstra, T. (2011). Towards an artificial therapy assistant: Measuring Excessive Stress from Speech. In V. Traver, A. Fred, J. Filipe, and H. Gamboa (Eds.), *Health Informatics 2011: Proceedings of the International Conference on Health Informatics*, p. 357–363. January 26-29, Rome – Italy.

25. **Broek, E.L. van den**, Janssen, J.H., Zwaag, M.D. van der, Westerink, J.H.D.M., & Healey, J.A. (2011). Prerequisites for Affective Signal Processing (ASP) – Part V: A response to comments and suggestions. In F. Babiloni, A. Fred, J. Filipe, and H. Gamboa (Eds.), *Biosignals 2011: Proceedings of the International Conference on Bio-Inspired Systems and Signal Processing*, p. 301–306. January 26-29, Rome – Italy.

24. **Broek, E.L. van den**, Nijholt, A., & Westerink, J.H.D.M. (2010). Unveiling Affective Signals. In E. Barakova, B. de Ruyter, and A.J. Spink (Eds.), *ACM Proceedings of Measuring Behavior 2010: Selected papers from the 7th international conference on methods and techniques in behavioral research*, Article No. a6. August 24–27, Eindhoven – The Netherlands.

23. **Broek, E.L. van den**, Nijholt, A., & Westerink, J.H.D.M. (2010). Unveiling Affective Signals. In A.J. Spink, F. Grieco, O.E. Krips, L.W.S. Loijens, L.P.J.J. Noldus, & P.H. Zimmerman (Eds.), *Proceedings of Measuring Behavior 2010: The 7th International Conference on Methods and Techniques in Behavioral Research*, p. 95–98. August 24–27, Eindhoven – The Netherlands.

22. **Broek, E.L. van den**, Nijholt, A., & Westerink, J.H.D.M. (2010). Symposium: Unveiling Affective Signals. In A.J. Spink, F. Grieco, O.E. Krips, L.W.S. Loijens, L.P.J.J. Noldus, &

P.H. Zimmerman (Eds.), *ACM Proceedings of Measuring Behavior 2010: The 7th International Conference on Methods and Techniques in Behavioral Research*, p. 94. August 24–27, Eindhoven – The Netherlands.

21. Zwaag, M.D. van der, **Broek, E.L. van den**, & Janssen, J.H. (2010). Guidelines for biosignal-driven HCI. In A. Girouard, R. Mandryk, L. Nacke, E.T. Solovey, D. Tan, & R.J.K. Jacob (Eds.), *Proceedings of ACM CHI2010 workshop - Brain, Body, and Bytes: Physiological user interaction*, p. 77–80. April 11, Atlanta, GA, USA.

20. **Broek, E.L. van den**, Zwaag, M.D. van der, Healey, J.A., Janssen, J.H., & Westerink, J.H.D.M. (2010). Prerequisites for Affective Signal Processing (ASP) - Part IV. In J. Kim & P. Karjalainen (Eds.), *Proceedings of the 1st International Workshop on Bio-inspired Human-Machine Interfaces and Healthcare Applications – `B-Interface 2010`*, p. 59–66. January 21, Valencia - Spain.

19. Schut, M.H., Tuinenbreijer, K., **Broek, E.L. van den**, Westerink, J.H.D.M. (2010). Biometrics for Emotion Detection (BED): Exploring the combination of Speech and ECG. In J. Kim & P. Karjalainen (Eds.), *Proceedings of the 1st International Workshop on Bio-inspired Human-Machine Interfaces and Healthcare Applications – `B-Interface 2010`*, p. 51–58. January 21, Valencia - Spain.

18. **Broek, E.L. van den**, Janssen, J.H., Healey, J.A., & Zwaag, M.D. van der. (2010). Prerequisites for Affective Signal Processing (ASP) - Part II. In A. Fred, J. Filipe & H. Gamboa (Eds.), *Biosignals 2010: Proceedings of the Third International Conference on Bio-Inspired Systems and Signal Processing*, p. 188–193. January 20-23, Valencia - Spain.

17. Sluis, F. van der, **Broek, E.L. van den**, & Dijkstra, T. (2010). Towards semi-automated assistance for the treatment of stress disorders. In A. Fred, J. Filipe & H. Gamboa (Eds.), *HealthInf 2010: Proceedings of the Third International Conference on Health Informatics*, p. 446–449. January 20-23, Valencia - Spain.

16. **Broek, E.L. van den**, Janssen, J.H., & Westerink, J.D.H.M. (2009). Guidelines for Affective Signal Processing (ASP): From lab to life. In J. Cohn, A. Nijholt, & M. Pantic (Eds.), *Proceedings of the IEEE International Conference on Affective Computing & Intelligent Interaction,* Vol. I, p. 704–709. September 10-12, Amsterdam - The Netherlands. *[special session introduction]*

15. Janssen, J.H., **Broek, E.L. van den**, & Westerink, J.D.H.M. (2009). Personalized affective music player. In J. Cohn, A. Nijholt, & M. Pantic (Eds.), *Proceedings of the IEEE International Conference on Affective Computing & Intelligent Interaction,* Vol. I, p. 472–477. September 10-12, Amsterdam - The Netherlands.

14. Zwaag, M.D. van der, Westerink, J.D.H.M., & **Broek, E.L. van den** (2009). Deploying music characteristics for an affective music player. In J. Cohn, A. Nijholt, & M. Pantic (Eds.), *Proceedings of the IEEE International Conference on Affective Computing & Intelligent*

Interaction, Vol. I, p. 459–465. September 10-12, Amsterdam - The Netherlands.

13. **Broek, E.L. van den**, Sluis, F. van der, & Dijkstra, T. (2009). Therapy Progress Indicator: Combining speech parameters and the subjective unit of distress. In J. Cohn, A. Nijholt, & M. Pantic (Eds.), *Proceedings of the IEEE International Conference on Affective Computing & Intelligent Interaction,* Vol. I, 381-386. September 10-12, Amsterdam - The Netherlands.

12. Janssen, J.H. & **Broek, E.L. van den** (2009). Guidelines for Mobile Emotion Measurement. In A. Geven, M. Tscheligi & L. Noldus (Eds.), *Proceedings of the ACM MobileHCI workshop – Measuring Mobile Emotions: Measuring the Impossible?*, p. 17–20. September 15, Bonn - Germany.

11. Bialoskorski, L.S.S., Westerink, J.H.D.M., & **Broek, E.L. van den** (2009). Mood Swings: An Affective Interactive Art System. *Lecture Notes of ICST, 9*, 181–186.

10. Bialoskorski, L.S.S., Westerink, J.H.D.M., & **Broek, E.L. van den** (2009). Come, see and experience affective interactive art. In A. Nijholt, D. Reidsma, and H. Hondorp (Eds.), *Program and demonstration papers of the Third International Conference on Intelligent Technologies for Interactive Entertainment (INTETAIN)*, 1–3. June 22-24, Amsterdam - The Netherlands.

9. **Broek, E.L. van den**, Janssen, J.H., Westerink, J.H.D.M., & Healey, J.A. (2009). Prerequisites for Affective Signal Processing (ASP). In P. Encarnação and A. Veloso (Eds.), Biosignals 2009: Proceedings of the International Conference on Bio-Inspired Systems and Signal Processing, p. 426–433. January 14-17, Porto - Portugal.

8. **Broek, E.L. van den**, Lisý, V., Westerink, J.D.H.M., Schut, M.H., & Tuinenbreijer, K. (2009). Biosignals as an Advanced Man-Machine Interface. In P. Encarnação and A. Veloso (Eds.), Biosignals 2009: Proceedings of the International Conference on Bio-Inspired Systems and Signal Processing, p. IS15–IS24. January 14-17, Porto - Portugal. *[invited]*

7. **Broek, E.L. van den**, Lisý, V., Westerink, J.D.H.M., Schut, M.H., & Tuinenbreijer, K. (2009). Biosignals as an Advanced Man-Machine Interface. In T.F.B. Filho and H. Gamboa (Eds.), Biodevices 2009: Proceedings of the International Conference on Biomedical Electronics and Devices, p. IS15–IS24. January 14-17, Porto - Portugal. *[invited]*

6. **Broek, E.L. van den**, Lisý, V., Westerink, J.D.H.M., Schut, M.H., & Tuinenbreijer, K. (2009). Biosignals as an Advanced Man-Machine Interface. In L. Azevedo and A.R. Londral (Eds.), Healthinf 2009: Proceedings of the International Conference on Health Informatics, p. IS15–IS24. January 14-17, Porto - Portugal. *[invited]*

5. Bialoskorski, L.S.S., Westerink, J.H.D.M., and **Broek, E.L. van den** (2008). Introducing Mood Swings. In P. Markopoulos, J. Hoonhout, I. Soute, and J. Read (Eds), Interna-

tional Conference on Fun and Games: Extended abstracts - Posters, demonstrations, and workshops, p. 50–53. October 20-21, Eindhoven - The Netherlands.

4. Drunen, A., **Broek, E.L. van den**, and Heffelaar, T. (2008). uLog: Towards attention aware user-system interactions measurement. In A. J. Spink and M. R. Ballintijn and N. D. Bogers and F. Grieco and L. W. S. Loijens and L. P. J. J. Noldus and G. Smit and P. H. Zimmerman (Eds.), Proceedings of Measuring Behavior 2008: 6th International Conference on Methods and Techniques in Behavioral Research, p. 109–110. August 26-29, Maastricht - The Netherlands.

3. **Broek, E.L. van den**, Schut, M.H., Westerink, J.H.D.M., Herk, J. van, and Tuinenbreijer, K. (2006). Computing emotion awareness through facial electromyography. *Lecture Notes in Computer Science (Human-Computer Interaction), 3979*, 51–62.

2. **Broek, E.L. van den**, Schut, M.H., Tuinenbreijer, K., and Westerink, J.H.D.M. (2006). Communication and Persuasion Technology: Psychophysiology of Emotions and User-Profiling. *Lecture Notes in Computer Science (Persuasive Technology), 3962*, 154–157.

1. **Broek, E.L. van den** (2005). Empathic Agent Technology (EAT). In L. Johnson, D. Richards, E. Sklar, and U. Wilensky (Eds.), *Proceedings of the AAMAS–05 Agent-Based Systems for Human Learning (ABSHL) workshop*, p. 59–67. July 26, Utrecht – The Netherlands.

Patents

3. Janssen, J.H., Westerink, J.H.D.M., & **Broek, E.L. van den** (2009). Method and system for selecting items using physiological parameters. International Patent Application No. *[processed]* ID: 673745 (PH013128EP1), filed on April 02, 2009.

2. Bialoskorski, L.S.S., Westerink, J.H.D.M., & **Broek, E.L. van den** (2009). System for sensing emotions in movement and expressing them with light or sound (MoodSwings) International Patent Application No. *[processed]* ID: 674616 (PH012766EP1), filed on March 10, 2009.

1. Westerink, J.H.D.M., **Broek, E.L. van den**, Schut, M.H., Tuinenbreijer, K. (2007). Higher order GSR-measurement interpretation indicating emotions. International Patent Application No. PCT/IB2008/050477 (PH007322), filed on February 11, 2007.

SIKS Dissertation Series

Since 1998, all dissertations written by Ph.D.-students who have conducted their research under auspices of a senior research fellow of the SIKS research school are published in the SIKS Dissertation Series. This monograph is the **328**th in the series.

2011-30 Egon L. van den Broek (UT), *Affective Signal Processing (ASP): Unraveling the mystery of emotions*

2011-29 Faisal Kamiran (TUE), *Discrimination-aware Classification*

2011-28 Rianne Kaptein (UVA), *Effective Focused Retrieval by Exploiting Query Context and Document Structure*

2011-27 Aniel Bhulai (VU), *Dynamic website optimization through autonomous management of design patterns*

2011-26 Matthijs Aart Pontier (VU), *Virtual Agents for Human Communication - Emotion Regulation and Involvement-Distance Trade-Offs in Embodied Conversational Agents and Robots*

2011-25 Syed Waqar ul Qounain Jaffry (VU), *Analysis and Validation of Models for Trust Dynamics*

2011-24 Herwin van Welbergen (UT), *Behavior Generation for Interpersonal Coordination with Virtual Humans On Specifying, Scheduling and Realizing Multimodal Virtual Human Behavior*

2011-23 Wouter Weerkamp (UVA), *Finding People and their Utterances in Social Media*

2011-22 Junte Zhang (UVA), *System Evaluation of Archival Description and Access*

2011-21 Linda Terlouw (TUD), *Modularization and Specification of Service-Oriented Systems*

2011-20 Qing Gu (VU), *Guiding service-oriented software engineering - A view-based approach*

2011-19 Ellen Rusman (OU), *The Mind's Eye on Personal Profiles*

2011-18 Mark Ponsen (UM), *Strategic Decision-Making in complex games*

2011-17 Jiyin He (UVA), *Exploring Topic Structure: Coherence, Diversity and Relatedness*

2011-16 Maarten Schadd (UM), *Selective Search in Games of Different Complexity*

2011-15 Marijn Koolen (UvA), *The Meaning of Structure: the Value of Link Evidence for Information Retrieval*

2011-14 Milan Lovric (EUR), *Behavioral Finance and Agent-Based Artificial Markets*

2011-13 Xiaoyu Mao (UvT), *Airport under Control. Multiagent Scheduling for Airport Ground Handling*

2011-12 Carmen Bratosin (TUE), *Grid Architecture for Distributed Process Mining*

2011-11 Dhaval Vyas (UT), *Designing for Awareness: An Experience-focused HCI Perspective*

2011-10 Bart Bogaert (UvT), *Cloud Content Contention*

2011-09 Tim de Jong (OU), *Contextualised Mobile Media for Learning*

2011-08 Nieske Vergunst (UU), *BDI-based Generation of Robust Task-Oriented Dialogues*

2011-07 Yujia Cao (UT), *Multimodal Information Presentation for High-Load Human Computer Interaction*

2011-06 Yiwen Wang (TUE), *Semantically-Enhanced Recommendations in Cultural Heritage*

2011-05 Base van der Raadt (VU), *Enterprise Architecture Coming of Age - Increasing the Performance of an Emerging Discipline*

2011-04 Hado van Hasselt (UU), *Insights in Reinforcement Learning Formal analysis and empirical evaluation of temporal-difference learning algorithms*

2011-03 Jan Martijn van der Werf (TUE), *Compositional Design and Verification of Component-Based Information Systems*

2011-02 Nick Tinnemeier (UU), *Organizing Agent Organizations. Syntax and Operational Semantics of an Organization-Oriented Programming Language*

2011-01 Botond Cseke (RUN), *Variational Algorithms for Bayesian Inference in Latent Gaussian Models*

2010-53 Edgar Meij (UVA), *Combining Concepts and Language Models for Information Access*

2010-52 Peter-Paul van Maanen (VU), *Adaptive*

Support for Human-Computer Teams: Exploring the Use of Cognitive Models of Trust and Attention

2010-51 Alia Khairia Amin (CWI), *Understanding and supporting information seeking tasks in multiple sources*

2010-50 Bouke Huurnink (UVA), *Search in Audiovisual Broadcast Archives*

2010-49 Jahn-Takeshi Saito (UM), *Solving difficult game positions*

2010-48 Milan Lovric (EUR), *Behavioral Finance and Agent-Based Artificial Markets*

2010-47 Chen Li (UT), *Mining Process Model Variants: Challenges, Techniques, Examples*

2010-46 Vincent Pijpers (VU), *e3alignment: Exploring Inter-Organizational Business-ICT Alignment*

2010-45 Vasilios Andrikopoulos (UvT), *A theory and model for the evolution of software services*

2010-44 Pieter Bellekens (TUE), *An Approach towards Context-sensitive and User-adapted Access to Heterogeneous Data Sources, Illustrated in the Television Domain*

2010-43 Peter van Kranenburg (UU), *A Computational Approach to Content-Based Retrieval of Folk Song Melodies*

2010-42 Sybren de Kinderen (VU), *Needs-driven service bundling in a multi-supplier setting - the computational e3-service approach*

2010-41 Guillaume Chaslot (UM), *Monte-Carlo Tree Search*

2010-40 Mark van Assem (VU), *Converting and Integrating Vocabularies for the Semantic Web*

2010-39 Ghazanfar Farooq Siddiqui (VU), *Integrative modeling of emotions in virtual agents*

2010-38 Dirk Fahland (TUE), *From Scenarios to components*

2010-37 Niels Lohmann (TUE), *Correctness of services and their composition*

2010-36 Jose Janssen (OU), *Paving the Way for Lifelong Learning; Facilitating competence development through a learning path specification*

2010-35 Dolf Trieschnigg (UT), *Proof of Concept: Concept-based Biomedical Information Retrieval*

2010-34 Teduh Dirgahayu (UT), *Interaction Design in Service Compositions*

2010-33 Robin Aly (UT), *Modeling Representation Uncertainty in Concept-Based Multimedia Retrieval*

2010-32 Marcel Hiel (UvT), *An Adaptive Service Oriented Architecture: Automatically solving Interoperability Problems*

2010-31 Victor de Boer (UVA), *Ontology Enrichment from Heterogeneous Sources on the Web*

2010-30 Marieke van Erp (UvT), *Accessing Natural History - Discoveries in data cleaning, structuring, and retrieval*

2010-29 Stratos Idreos (CWI), *Database Cracking: Towards Auto-tuning Database Kernels*

2010-28 Arne Koopman (UU), *Characteristic Relational Patterns*

2010-27 Marten Voulon (UL), *Automatisch contracteren*

2010-26 Ying Zhang (CWI), *XRPC: Efficient Distributed Query Processing on Heterogeneous XQuery Engines*

2010-25 Zulfiqar Ali Memon (VU), *Modelling Human-Awareness for Ambient Agents: A Human Mindreading Perspective*

2010-24 Dmytro Tykhonov (TUD), *Designing Generic and Efficient Negotiation Strategies*

2010-23 Bas Steunebrink (UU), *The Logical Structure of Emotions*

2010-22 Michiel Hildebrand (CWI), *End-user Support for Access to Heterogeneous Linked Data*

2010-21 Harold van Heerde (UT), *Privacy-aware data management by means of data degradation*

2010-20 Ivo Swartjes (UT), *Whose Story Is It Anyway? How Improv Informs Agency and Authorship of Emergent Narrative*

2010-19 Henriette Cramer (UvA), *People's Responses to Autonomous and Adaptive Systems*

2010-18 Charlotte Gerritsen (VU), *Caught in the Act: Investigating Crime by Agent-Based Simulation*

2010-17 Spyros Kotoulas (VU), *Scalable Discovery of Networked Resources: Algorithms, Infrastructure, Applications*

2010-16 Sicco Verwer (TUD), *Efficient Identification of Timed Automata, theory and practice*

2010-15 Lianne Bodenstaff (UT), *Managing Dependency Relations in Inter-Organizational Models*

2010-14 Sander van Splunter (VU), *Automated Web Service Reconfiguration*

2010-13 Gianluigi Folino (RUN), *High Performance Data Mining using Bio-inspired techniques*

2010-12 Susan van den Braak (UU), *Sensemaking software for crime analysis*

2010-11 Adriaan Ter Mors (TUD), *The world according to MARP: Multi-Agent Route Planning*

2010-10 Rebecca Ong (UL), *Mobile Communication and Protection of Children*

2010-09 Hugo Kielman (UL), *A Politiele gegevensverwerking en Privacy, Naar een effectieve waarborging*

2010-08 Krzysztof Siewicz (UL), *Towards an Improved Regulatory Framework of Free Software. Protecting user freedoms in a world of software communities and eGovernments*

2010-07 Wim Fikkert (UT), *Gesture Interaction at a Distance*

2010-06 Sander Bakkes (UvT), *Rapid Adaptation of Video Game AI*

2010-05 Claudia Hauff (UT), *Predicting the Effectiveness of Queries and Retrieval Systems*

2010-04 Olga Kulyk (UT), *Do You Know What I Know? Situational Awareness of Co-located Teams in Multidisplay Environments*

2010-03 Joost Geurts (CWI), *A Document Engineering Model and Processing Framework for Multimedia documents*

2010-02 Ingo Wassink (UT), *Work flows in Life Science*

2010-01 Matthijs van Leeuwen (UU), *Patterns that Matter*

2009-46 Loredana Afanasiev (UvA), *Querying XML: Benchmarks and Recursion*

2009-45 Jilles Vreeken (UU), *Making Pattern Mining Useful*

2009-44 Roberto Santana Tapia (UT), *Assessing Business-IT Alignment in Networked Organizations*

2009-43 Virginia Nunes Leal Franqueira (UT), *Finding Multi-step Attacks in Computer Networks using Heuristic Search and Mobile Ambients*

2009-42 Toine Bogers, *Recommender Systems for Social Bookmarking*

2009-41 Igor Berezhnyy (UvT), *Digital Analysis of Paintings*

2009-40 Stephan Raaijmakers (UvT), *Multinomial Language Learning: Investigations into the Geometry of Language*

2009-39 Christian Stahl (TUE, Humboldt–Universitaet zu Berlin), *Service Substitution – A Behavioral Approach Based on Petri Nets*

2009-38 Riina Vuorikari (OU), *Tags and self-organisation: a metadata ecology for learning resources in a multilingual context*

2009-37 Hendrik Drachsler (OUN), *Navigation Support for Learners in Informal Learning Networks*

2009-36 Marco Kalz (OUN), *Placement Support for Learners in Learning Networks*

2009-35 Wouter Koelewijn (UL), *Privacy en Politiegegevens; Over geautomatiseerde normatieve informatie-uitwisseling*

2009-34 Inge van de Weerd (UU), *Advancing in Software Product Management: An Incremental Method Engineering Approach*

2009-33 Khiet Truong (UT), *How Does Real Affect Affect Affect Recognition In Speech?*

2009-32 Rik Farenhorst (VU) and Remco de Boer (VU), *Architectural Knowledge Management: Supporting Architects and Auditors*

2009-31 Sofiya Katrenko (UVA), *A Closer Look at Learning Relations from Text*

2009-30 Marcin Zukowski (CWI), *Balancing vectorized query execution with bandwidth-optimized storage*

2009-29 Stanislav Pokraev (UT), *Model-Driven Semantic Integration of Service-Oriented Applications*

2009-28 Sander Evers (UT), *Sensor Data Management with Probabilistic Models*

2009-27 Christian Glahn (OU), *Contextual Support of social Engagement and Reflection on the Web*

2009-26 Fernando Koch (UU), *An Agent-Based Model for the Development of Intelligent Mobile Services*

2009-25 Alex van Ballegooij (CWI), *RAM: Array Database Management through Relational Mapping*

2009-24 Annerieke Heuvelink (VUA), *Cognitive Models for Training Simulations*

2009-23 Peter Hofgesang (VU), *Modelling Web Usage in a Changing Environment*

2009-22 Pavel Serdyukov (UT), *Search For Expertise: Going beyond direct evidence*

2009-21 Stijn Vanderlooy (UM), *Ranking and Reliable Classification*

2009-20 Bob van der Vecht (UU), *Adjustable Autonomy: Controling Influences on Decision Making*

2009-19 Valentin Robu (CWI), *Modeling Preferences, Strategic Reasoning and Collaboration in Agent-Mediated Electronic Markets*

2009-18 Fabian Groffen (CWI), *Armada, An Evolving Database System*

2009-17 Laurens van der Maaten (UvT), *Feature*

Extraction from Visual Data

2009-16 Fritz Reul (UvT), *New Architectures in Computer Chess*

2009-15 Rinke Hoekstra (UVA), *Ontology Representation - Design Patterns and Ontologies that Make Sense*

2009-14 Maksym Korotkiy (VU), *From ontology-enabled services to service-enabled ontologies (making ontologies work in e-science with ONTO-SOA)*

2009-13 Steven de Jong (UM), *Fairness in Multi-Agent Systems*

2009-12 Peter Massuthe (TUE), Humboldt–Universitaet zu Berlin), *perating Guidelines for Services*

2009-11 Alexander Boer (UVA), *Legal Theory, Sources of Law and the Semantic Web*

2009-10 Jan Wielemaker (UVA), *Logic programming for knowledge-intensive interactive applications*

2009-09 Benjamin Kanagwa (RUN), *Design, Discovery and Construction of Service-oriented Systems*

2009-08 Volker Nannen (VU), *Evolutionary Agent-Based Policy Analysis in Dynamic Environments*

2009-07 Ronald Poppe (UT), *Discriminative Vision-Based Recovery and Recognition of Human Motion*

2009-06 Muhammad Subianto (UU), *Understanding Classification*

2009-05 Sietse Overbeek (RUN), *Bridging Supply and Demand for Knowledge Intensive Tasks - Based on Knowledge, Cognition, and Quality*

2009-04 Josephine Nabukenya (RUN), *Improving the Quality of Organisational Policy Making using Collaboration Engineering*

2009-03 Hans Stol (UvT), *A Framework for Evidence-based Policy Making Using IT*

2009-02 Willem Robert van Hage (VU), *Evaluating Ontology-Alignment Techniques*

2009-01 Rasa Jurgelenaite (RUN), *Symmetric Causal Independence Models*